



XLIX Konferencja

Statystyka Matematyczna

Chęciny, 9–12 grudnia 2024 r.



Wydział Matematyki
i Nauk Informatycznych

Politechnika Warszawska

Organizatorzy

Wydział Matematyki i Nauk Informatycznych Politechniki Warszawskiej

Wydział Matematyki, Informatyki i Mechaniki Uniwersytetu Warszawskiego

Polskie Towarzystwo Matematyczne

Komitety organizacyjny

Przemysław Grzegorzewski (przewodniczący)

Błażej Miasojedow

Jan Mielniczuk

Jacek Miękiś

Agnieszka Piliszek

Krzysztof Rudaś

Paweł Teisseyre

Przeszło pięćdziesiąt lat temu, a dokładniej w dniach 6–12 grudnia 1973, odbyła się w Wiśle pierwsza konferencja „Statystyka matematyczna”, zorganizowana przez Komisję ds. Rozwoju Statystyki Matematycznej w Polsce, powołaną rok wcześniej przez Komitet Nauk Matematycznych Polskiej Akademii Nauk. Inicjatorem powstania Komisji był prof. Tadeusz Caliński, a jej pierwszym przewodniczącym prof. Józef Łukaszewicz (więcej informacji o historii i bieżącej działalności Komisji Statystyki oraz osobach, które ją tworzyły, można znaleźć na stronie internetowej <https://www.ibspan.waw.pl/komisja.statystyki/>).

Choć w ciągu minionego półwiecza zmieniały się nazwy i struktury wspomnianych wyżej ciał, organizowane niemal corocznie konferencje „Statystyka matematyczna” w miarę upływu lat zyskiwały na znaczeniu, przyciągając uczestników nie tylko z Polski, ale również z zagranicy. Sprzyjał temu wysoki poziom merytoryczny referatów, wykłady zaproszonych gości, w tym wybitnych statystyków o międzynarodowej renomie oraz konkurs na najlepszą prezentację dla młodych statystyków.

W latach 1973-2022 zorganizowanych zostało 48 konferencji ze statystyki matematycznej, w tym kilkanaście międzynarodowych (pełną listę konferencji można znaleźć na stronie <https://sm2024.mini.pw.edu.pl/node/30>). Konferencje te organizowane były przez różne krajowe ośrodki naukowe. Ponieważ większość z nich odbyła się w Wiśle, lokalizacja ta stała się w środowisku naukowym synonimem konferencji „Statystyczna matematyczna”.

Z perspektywy minionych lat widać, iż konferencje z cyklu „Statystyka matematyczna” stały się unikatowym wydarzeniem naukowym, niezmiernie ważnym dla środowiska zajmującego się statystyką matematyczną w naszym kraju – zarówno dla uczonych pracujących twórczo nad różnymi problemami statystyki, dla osób stosujących metody statystyczne w praktyce jak i dla doktorantów oraz młodych pracowników nauki zainteresowanych tą dyscypliną. Bez względu na lokalizację, organizacja „wiślańskich” konferencji stała się tradycją symbolizującą ciągłość i jedność środowiska statystyków w Polsce.

Tegoroczna, XLIX Konferencja „Statystyka Matematyczna 2024” odbędzie się w dniach 9-12 grudnia 2024 r. w Europejskim Centrum Edukacji Geologicznej w Chęcinach (<https://www.eceg.uw.edu.pl>). Jej organizatorami są:

- Wydział Matematyki i Nauk Informatycznych Politechniki Warszawskiej
- Wydział Matematyki, Informatyki i Mechaniki Uniwersytetu Warszawskiego
- Polskie Towarzystwo Matematyczne.

Życzymy owocnego uczestnictwa w konferencji

W imieniu Komitetu Organizacyjnego
– *prof. dr hab. Przemysław Grzegorzewski*

Spis treści

Program konferencji	8
Zaproszeni wykładowcy	13
Jak badać metody analizy skupień?	14
<i>Marek Gągolewski</i>	
Topologia w statystyce, statystyka w topologii	16
<i>Paweł Dłotko</i>	
Streszczenia	18
Zastosowanie modelu AMMI do oceny odmian soi w północnej Polsce	19
<i>Konrad Banaś, Agnieszka Osiecka, Tomasz Lenartowicz, Agnieszka Łącka, Henryk Bujak, Marcin Przystalski</i>	
Charakteryzacje rozkładów prawdopodobieństwa przez regresje uogólnionych statystyk porządkowych	20
<i>Mariusz Bieniek</i>	
Estymacja dużych macierzy precyzji i identyfikacja modeli grafowych poprzez regularyzację macierzy korelacji warunkowych	21
<i>Małgorzata Bogdan, Adam Chojecki, Ivan Hejny, Bartosz Kołodziejek, Jonas Wallin</i>	
Szacowanie poziomów zanieczyszczeń środowiskowych	22
<i>Jacek Bojarski, Dawid Drapiński, Maciej Niedziela, Katarzyna Pypka, Jarosław Stein, Gabriel Zaborniak, Piotr Ziembicki, Kornel Żuber</i>	
Bayesian Model Selection for Colored Probabilistic Graphical Models	23
<i>Adam Przemysław Chojecki, Piort Graczyk, Hideyuki Ishi, Bartosz Kołodziejek</i>	
Wykrywanie struktur zależności dla pary zmiennych losowych	24
<i>Bogdan Ćmiel, Teresa Ledwina</i>	
Discrete parametric graphical models	25
<i>Iza Danielewska, Bartosz Kołodziejek, Jacek Wesółowski Xiaolin Zeng</i>	
Poprawki do praw Zipfa i Heapsa oparte na modelowaniu frakcji hapaksów	26
<i>Lukasz Dębowski</i>	
Optymalny wybór długości bloku bootstrapowego dla procesów okresowo skorelowanych	27
<i>Patrice Bertail, <u>Anna Dudek</u></i>	

Warunkowe estymatory największej wiarygodności w modelu wielowymiarowym	28
<i>Katarzyna Filipiak</i>	
Graficzny model współwystępowania chorób alergicznych	29
<i>Konrad Furmańczyk, Wojciech Niemirow, Mariola Chrzanowska, Marta Zalewska</i>	
Utracona pojemność ważonych systemów k-spośród-n z dyskretnymi czasami życia komponentów	30
<i>Agnieszka Goroncy, Krzysztof Jasiński, Faustyna Korejwo, Marta Rudzate</i>	
Estymacja przedziałowa efektu zabiegowego na podstawie danych nieostrych	31
<i>Przemysław Grzegorzewski</i>	
Grafove modele oddziaływań wyższych rzędów	32
<i>Aleksandra Grzelik, Mateusz Iskrzyński, Aleksandra Puchalska</i>	
Regularyzacja w regresji różnicowej i modelowaniu przyczynowym	33
<i>Szymon Jaroszewicz, Krzysztof Rudaś</i>	
O liczbie komponentów w poszczególnych stanach w trzystanowym systemie k-spośród-n	34
<i>Krzysztof Jasiński, Agnieszka Goroncy</i>	
Estymacja parametrów procesu Poissona z ograniczoną funkcją wartości średniej	35
<i>Alicja Jokiel-Rokita, Ryszard Magiera</i>	
Integracja różnych czynników doświadczalnych i grup cech w zaawansowanych modelach sieciowych roślin	36
<i>Marta Kańczurzevska-Staszak, Aneta Sawikowska</i>	
Pattern recovery with GSLOPE	37
<i>Bartosz Kołodziejek</i>	
Wpływ strategii wstępnego przetwarzania danych na jakość modeli opartych na drzewach w kontekście AutoML	38
<i>Hubert Ruczyński, Anna Kozak</i>	
Szeregi czasowe stacjonarne o własnościach cyklicznych	39
<i>Łukasz Lenart</i>	
Analiza spektralna i subsampling dla procesów spektralnie skorelowanych	40
<i>Anna E. Dudek, Bartosz Majewski</i>	
Bootstrap for high-dimensional Markov chains	41
<i>Patrice Bertail, Anna Dudek, Karolina Marek</i>	
Testy statystyczne dla indeksu Jaccarda	42
<i>Błażej Miasojedow</i>	
Weryfikacja założenia SCAR (Selected Completely at Random) dla danych typu PU (Positive Unlabeled)	43
<i>P. Teisseyre, K. Furmańczyk, J. Mielniczuk</i>	
Likelihood ratio test for covariance matrix under the multivariate and matrix-variate t distribution	44
<i>Katarzyna Filipiak, Daniel Klein, Stepan Mazur, Malwina Mrowińska</i>	

Porównanie metod modelowania zbiorów szeregów czasowych i oceniania ich skuteczności	45
<i>Karol Opara</i>	
Wpływ struktury systemu na przenoszenie własności starzenia jego komponentów	46
<i>Jagoda Papis, Tomasz Rychlik, Magdalena Szymkowiak</i>	
Estimation of conditional inequality measures	47
<i>Alicja Jokiel-Rokita, Sylwester Piątek, Rafał Topolnicki</i>	
Analiza czasu przeżycia pacjentów z ostrym zespołem wieńcowym	48
<i>Katarzyna Pypka, Magdalena Wojciech, Konrad Pieszko</i>	
Łączna estymacja prawdopodobieństwa a posteriori i funkcji skłonności w modelu PU	49
<i>Jan Mielniczuk, Wojciech Rejchel, Paweł Teisseyre</i>	
Zamiana wartości klas w modelowaniu przyczynowości dla niezbalansowanych zrandomizowanych danych	50
<i>Krzysztof Rudaś, Szymon Jaroszewicz</i>	
Autocovariance estimation for generalized almost-cyclostationary processes with unknown lag-dependent cycle frequencies	51
<i>Anna Dudek, Antonio Napolitano, Jakub Rutkowski, Jakub Wojdyła</i>	
Warunki skończoności i oszacowania momentów uogólnionych statystyk pozycyjnych	52
<i>Mariusz Bieniek, Tomasz Rychlik</i>	
Zastosowanie metod wielowymiarowych w badaniu wpływu makuchu lniatego na bazylię w warunkach suszy	53
<i>Aneta Sawikowska, Marta Kańczurzevska-Staszak</i>	
Analiza układów wierszowo-kolumnowych – model losowy	54
<i>Idzi Siatkowski</i>	
Funkcjonalna analiza wariancji w detekcji stanu technicznego maszyn pracujących przy zmiennym obciążeniu	55
<i>Maria Skupień</i>	
Rozkład normalny skośny ze składnikiem uplastyczniającym II	56
<i>Damian Stoltmann</i>	
Skew Bimodal Power Laplace Distribution	57
<i>Piotr Sulewski</i>	
O optymalnej adaptacji w poissonowskich problemach odwrotnych	58
<i>Zbigniew Szkutnik</i>	
Kluczowe rozkłady życia komponentów gwarantujące systemowi dziedziczenie wybranych własności starzenia	59
<i>Jorge Navarro, Tomasz Rychlik, Magdalena Szymkowiak</i>	
Metody klasyfikacji w sytuacji przesunięcia rozkładu warunkowego zmiennej odpowiedzi między zbiorem treningowym i testowym	60
<i>Paweł Teisseyre, Jan Mielniczuk</i>	

Model klasyfikacyjny dla spraw sądowych Banku	61
<i>Łukasz Waszak</i>	
Semi-supervised learning with semi-continuous hidden Markov models . .	62
<i>Filip Wichrowski, Katarzyna Kaczmarek-Majer</i>	
Test permutacyjny w problemie k-prób dla danych prawostronnie cenzurowanych	63
<i>Grzegorz Wyłupek</i>	
Przedziały ufności dla „wrażliwej proporcji” w badaniach ankietowych . .	64
<i>Marta Zalewska, Wojciech Niemirowicz</i>	
Lista uczestników	65
Indeks	70

Program konferencji

Poniedziałek, 9 grudnia 2024

8:00 – 9:00	ŚNIADANIE
9:15 – 9:30	OTWARCIE KONFERENCJI
9:30 – 10:30	Marek Gągolewski <i>Jak badać metody analizy skupień? (część I)</i>
10:30 – 10:50	PRZERWA NA KAWĘ
10:50 – 11:10	Marta Zalewska <i>Przedziały ufności dla „wrażliwej proporcji” w badaniach ankietowych</i>
11:10 – 11:30	Błażej Miasojedow <i>Testy statystyczne dla indeksu Jaccarda</i>
11:30 – 11:50	Grzegorz Wyłupek <i>Test permutacyjny w problemie k-prób dla danych prawostronnie cenzurowanych</i>
11:50 – 12:10	Zbigniew Szkutnik <i>O optymalnej adaptacji w poissonowskich problemach odwrotnych</i>
12:10 – 12:40	PRZERWA NA KAWĘ
12:40 – 13:00	Jagoda Papis <i>Wpływ struktury systemu na przenoszenie własności starzenia jego komponentów</i>
13:00 – 13:20	Sylwester Piątek <i>Estimation of conditional inequality measures</i>
13:20 – 13:40	Malwina Mrowińska <i>Likelihood ratio test for covariance matrix under the multivariate and matrix-variate t distribution</i>
14:00 – 15:00	OBIAD
16:00 – 16:20	Filip Wichrowski <i>Semi-supervised learning with semi-continuous hidden Markov models</i>
16:20 – 16:40	Karolina Marek <i>Bootstrap for high-dimensional Markov chains</i>
16:40 – 17:00	Jakub Rutkowski <i>Autocovariance estimation for generalized almost-cyclostationary processes with unknown lag-dependent cycle frequencies</i>
17:00 – 17:20	Bartosz Majewski <i>Analiza spektralna i subsampling dla procesów spektralnie skorelowanych</i>
17:20 – 17:50	PRZERWA NA KAWĘ
17:50 – 18:10	Anna Kozak <i>Wpływ strategii wstępnego przetwarzania danych na jakość modeli opartych na drzewach w kontekście AutoML</i>
18:10 – 18:30	Łukasz Dębowski <i>Poprawki do praw Zipfa i Heapsa oparte na modelowaniu frakcji hapaksów</i>
18:30 – 18:50	Jacek Bojarski <i>Szacowanie poziomów zanieczyszczeń środowiskowych</i>
19:00 – 20:00	KOLACJA

Wtorek, 10 grudnia 2024

8:00 – 9:00	ŚNIADANIE
9:30 – 10:30	Marek Gągolewski <i>Jak badać metody analizy skupień? (część II)</i>
10:30 – 10:50	PRZERWA NA KAWĘ
10:50 – 11:10	Bogdan Ćmiel <i>Wykrywanie struktur zależności dla pary zmiennych losowych</i>
11:10 – 11:30	Konrad Furmańczyk <i>Graficzny model współwystępowania chorób alergicznych</i>
11:30 – 11:50	Bartosz Kołodziejek <i>Pattern recovery with GSLOPE</i>
11:50 – 12:10	Małgorzata Bogdan <i>Estymacja dużych macierzy precyzji i identyfikacja modeli grafowych poprzez regularyzację macierzy korelacji warunkowych</i>
12:10 – 12:40	PRZERWA NA KAWĘ
12:40 – 13:00	Aleksandra Grzelik <i>Grafowe modele oddziaływań wyższych rzędów</i>
13:00 – 13:20	Iza Danielewska <i>Discrete parametric graphical models</i>
13:20 – 13:40	Adam Chojecki <i>Bayesian Model Selection for Colored Probabilistic Graphical Models</i>
14:00 – 15:00	OBIAD
16:00 – 16:20	Konrad Banaś <i>Zastosowanie modelu AMMI do oceny odmian soi w północnej Polsce</i>
16:20 – 16:40	Damian Stoltmann <i>Rozkład normalny skośny ze składnikiem uplastyczniającym II</i>
16:40 – 17:00	Katarzyna Pypka <i>Analiza czasu przeżycia pacjentów z ostrym zespołem wieńcowym</i>
17:00 – 17:20	Maria Skupień <i>Funkcjonalna analiza wariancji w detekcji stanu technicznego maszyn pracujących przy zmiennym obciążeniu</i>
17:20 – 17:50	PRZERWA NA KAWĘ
17:50 – 18:10	Marta Kańczurzevska-Staszak <i>Integracja różnych czynników doświadczalnych i grup cech w zaawansowanych modelach sieciowych roślin</i>
18:10 – 18:30	Aneta Sawikowska <i>Zastosowanie metod wielowymiarowych w badaniu wpływu makuchu lnianego na bazylię w warunkach suszy</i>
18:30 – 18:50	Łukasz Waszak <i>Model klasyfikacyjny dla spraw sądowych Banku</i>
19:00 – 20:00	KOLACJA
20:00 – ...	ZEBRANIE KOMISJI STATYSTYKI

Środa, 11 grudnia 2024

8:00 – 9:00	ŚNIADANIE
9:30 – 10:30	Paweł Dłotko <i>Topologia w statystyce, statystyka w topologii (część I)</i>
10:30 – 10:50	PRZERWA NA KAWĘ
10:50 – 11:10	Krzysztof Jasiński <i>O liczbie komponentów w poszczególnych stanach w trzystanowym systemie k-spośród-n</i>
11:10 – 11:30	Agnieszka Goroncy <i>Utracona pojemność ważonych systemów k-spośród-n z dyskretnymi czasami życia komponentów</i>
11:30 – 11:50	Mariusz Bieniek <i>Charakteryzacje rozkładów prawdopodobieństwa przez regresje uogólnionych statystyk porządkowych</i>
11:50 – 12:10	Tomasz Rychlik <i>Warunki skończoności i oszacowania momentów uogólnionych statystyk pozycyjnych</i>
12:10 – 14:00	WYCIECZKA
14:00 – 15:00	OBIAD
16:00 – 16:20	Idzi Siatkowski <i>Analiza układów wierszowo-kolumnowych – model losowy</i>
16:20 – 16:40	Katarzyna Filipiak <i>Warunkowe estymatory największej wiarygodności w modelu wielowymiarowym</i>
16:40 – 17:00	Krzysztof Rudaś <i>Zamiana wartości klas w modelowaniu przyczynowości dla niezbalansowanych zrandomizowanych danych</i>
17:00 – 17:20	Szymon Jaroszewicz <i>Regularyzacja w regresji różnicowej i modelowaniu przyczynowy</i>
17:20 – 17:50	PRZERWA NA KAWĘ
17:50 – 18:10	Paweł Teisseyre <i>Metody klasyfikacji w sytuacji przesunięcia rozkładu warunkowego zmiennej odpowiedzi między zbiorem treningowym i testowym</i>
18:10 – 18:30	Wojciech Rejchel <i>Łączna estymacja prawdopodobieństwa a posteriori i funkcji skłonności w modelu PU</i>
18:30 – 18:50	Jan Mielniczuk <i>Weryfikacja założenia SCAR (Selected Completely at Random) dla danych typu PU (Positive Unlabeled)</i>
19:30 – ...	UROCZYSTA KOLACJA

Czwartek, 12 grudnia 2024

8:00 – 9:00	ŚNIADANIE
9:30 – 10:30	Paweł Dłotko <i>Topologia w statystyce, statystyka w topologii (część II)</i>
10:30 – 10:50	PRZERWA NA KAWĘ
10:50 – 11:10	Anna Dudek <i>Optymalny wybór długości bloku bootstrapowego dla procesów okresowo skorelowanych</i>
11:10 – 11:30	Łukasz Lenart <i>Szeregi czasowe stacjonarne o własnościach cyklicznych</i>
11:30 – 11:50	Alicja Jokiel-Rokita <i>Estymacja parametrów procesu Poissona z ograniczoną funkcją wartości średniej</i>
11:50 – 12:10	Karol Opara <i>Porównanie metod modelowania zbiorów szeregów czasowych i oceniania ich skuteczności</i>
12:10 – 12:40	PRZERWA NA KAWĘ
12:40 – 13:00	Piotr Sulewski <i>Skew Bimodal Power Laplace Distribution</i>
13:00 – 13:20	Magdalena Szymkowiak <i>Kluczowe rozkłady życia komponentów gwarantujące systemowi dziedziczenie wybranych własności starzenia</i>
13:20 – 13:40	Przemysław Grzegorzewski <i>Estymacja przedziałowa efektu zabiegowego na podstawie danych nieostrych</i>
14:00 – 15:00	OBIAD

Zaproszeni wykładowcy



Marek Gałolewski jest informatykiem pracującym na Wydziale Matematyki i Nauk Informatycznych Politechniki Warszawskiej. Absolwent tejże uczelni (2008 r.). Pełnomocnik Dziekana ds studiów w dziedzinie Inżynierii i Analizy Danych (Data Science) i jeden z głównych twórców tego kierunku. Doktorat i habilitację uzyskał w Instytucie Badań Systemowych PAN. W latach 2019–2014 pracował na Deakin University w Melbourne (Australia). Zajmuje się tworzeniem i badaniem metod analizy i modelowania danych. Redaktor działowy w czasopiśmie *Fuzzy Sets and Systems* (data science and aggregation). Autor, współautor lub redaktor ok. 100 publikacji, m.in. artykułów w czasopismach *Proceedings of the National Academy of Sciences (PNAS)*, *Journal of Statistical Software*, *R Journal*, *Journal of Classification*, *Information Fusion*, *International Journal of Forecasting*, *Statistical Modelling*, *Physica A: Statistical Mechanics and Its Applications*, *Information Sciences* i *Journal of Informetrics* oraz książek „Przetwarzanie i analiza danych w języku R”, „Deep R Programming” i „Minimalist Data Wrangling with Python”.

Jak badać metody analizy skupień?

Marek Gałolewski

Wydział Matematyki i Nauk Informatycznych, Politechnika Warszawska
Instytut Badań Systemowych PAN

Podręczniki szeroko pojętej statystycznej analizy danych mianują najczęściej analizę skupień (ang. *cluster analysis*) narzędziem służącym do grupowania obiektów w pewną liczbę niepustych i rozłącznych podzbiorów, „wewnątrz których obserwacje są sobie w jakimś określonym sensie bliskie” [1] lub „podobne” [2]. Interesuje nas tutaj „wykrycie w zbiorze danych tzw. »naturalnych« skupień, czyli skupień, które dają się w sensowny sposób interpretować” [2]. Tymczasem to, co rozumiemy przez „naturalne” skupienie, jest wciąż przedmiotem dyskusji (por. [3]).

Nie przeszkadza to jednak badaczom proponować coraz to nowsze algorytmy grupowania obiektów: od *Taksonomii Wrocławskiej* (*Single Linkage*) i innych metod hierarchicznych, przez metodę *K Średnich* i jej uogólnienia (np. algorytm *EM*), po metody spektralne, typu *DBSCAN* itp. Badamy je często w sposób znany z innych gałęzi uczenia maszynowego,

czyli sprawdzamy, czy w wyniku ich zastosowania na zbiorach referencyjnych otrzymujemy to, czego się spodziewaliśmy (np. [4]).

W niniejszym wykładzie rozważymy kilka problemów analizy porównawczej metod grupowania obserwacji, które omówiliśmy w pracach [5, 6, 7]:

- skąd wziąć dobre dane referencyjne (i dlaczego nie z bazy UCI),
- czy istnieje tylko jeden słuszny podział zbioru na skupienia,
- jak mierzyć podobieństwo podziałów,
- czy znane i powszechnie używane indeksy jakości grupowania rzeczywiście mierzą to, co obiecują?

Literatura

- [1] Koronacki J., Ćwik J., *Statystyczne systemy uczące się*, EXIT, Warszawa, 2008
- [2] Krzyśko M., Wołyński W., Górecki T., Skorzybut M., *Systemy uczące się*, WNT, Warszawa, 2008
- [3] Hennig C., What are the true clusters?, *Pattern Recognition Letters* **64** (2015), 53–62.
- [4] van Mechelen I., Boulesteix A.-L., Dangl R., Dean N., Hennig C., Leisch F., Steinley D., Warrens M.J., A white paper on good research practices in benchmarking: The case of cluster analysis, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* **13** (2023), e1511.
- [5] Gągolewski M., Normalised clustering accuracy: An asymmetric external cluster validity measure, *Journal of Classification* (2024), w druku, DOI:10.1007/s00357-024-09482-2.
- [6] Gągolewski M., A framework for benchmarking clustering algorithms, *SoftwareX* **20** (2022), 101270, <https://clustering-benchmarks.gagolewski.com/>.
- [7] Gągolewski M., Bartoszek M., Cena, A., Are cluster validity measures (in)valid?, *Information Sciences* **581** (2021), 620–636.



Paweł Dłotko jest matematykiem i informatykiem specjalizującym się w topologii stosowanej i obliczeniowej. Ukończył studia w Uniwersytecie Jagiellońskim, a doświadczenie badawcze zdobywał m.in. na Uniwersytecie Pensylwanii, w Inria Saclay oraz na Uniwersytecie w Swansea. Obecnie kieruje Centrum Dioscuri w Topologicznej Analizie Danych przy Instytucie Matematycznym Polskiej Akademii Nauk.

Jego prace naukowe koncentrują się na rozwijaniu metod topologicznej analizy danych i ich zastosowaniach w naukach stosowanych. Jest autorem narzędzi obliczeniowych, takich jak Persistence Landscape Toolbox, współtwórcą bibliotek Gudhi i RBallMapper, oraz współautorem licznych publikacji naukowych. W swojej działalności badawczej współpracuje z naukowcami z różnych dziedzin, takich jak inżynieria, biologia, medycyna czy ekonomia, przyczyniając się do integracji metod matematycznych w szerokim spektrum zastosowań.

Topologia w statystyce, statystyka w topologii

Paweł Dłotko

Instytut Matematyczny Polskiej Akademii Nauk

W proponowanej serii wykładów przedstawię przystępne wprowadzenie do metod topologicznej analizy danych (TDA) [1], ze szczególnym naciskiem na ich integrację ze statystyką. Metody TDA pozwalają w sposób ścisły opisać kształt rozważanych danych. Można ogólnie podzielić na dwie kategorie: ekstrakcję cech i wizualizację danych.

Pierwsza kategoria obejmuje wydobywanie z danych cech topologicznych. Mogą one być następnie wykorzystane w różnych procedurach statystycznych. Kluczowe metody w tej grupie to, między innymi, homologia persistentna [1, 2], krzywe oraz profile charakterystyki Eulera [3]. Wprowadzenie tych pojęć i pokazanie ich użyteczności w połączeniu ze standardowymi technikami statystycznymi będzie jednym z tematów wykładu. W szczególności, jako przykład użycia tych technik omówię TopoTesty [4], metodę wykorzystującą cechy topologiczne do rozwiązywania problemu testowania zgodności statystycznej.

Druga kategoria metod TDA koncentruje się na wizualizacji danych. Te metody dają analitykom wgląd w strukturę rozważanych danych ułatwiają wybór odpowiednich narzędzi do ich dalszej analizy. W tej kategorii znajdują się techniki takie jak Mapper [5], Ball Mapper [6] oraz niedawno opracowany ClusterGraph [7]. Zaprezentujemy ogólny zarys tych metod wraz z ich przykładowymi zastosowaniami.

Przedstawione koncepcje zostaną zilustrowane praktycznymi przykładami użycia kodu opracowanego w szczególności w Centrum Dioscuri w Analizie Danych Topologicznych [9], pokazujące praktyczne zastosowania tych metod TDA w kontekstach statystycznych. Popularnonaukowy wstęp do prezentowanych zagadnień można znaleźć w [2, 8]

Literatura

- [1] Herbert Edelsbrunner, John Harer, *Computational Topology: An Introduction*, American Mathematical Society, 2010.
- [2] Paweł Dłotko, *On the Shape That Matters – Topology and Geometry in Data Science*, EMS Magazine, No. 132, 2024.
- [3] Paweł Dłotko, Davide Gurnari, *Euler Characteristic Curves and Profiles: A Stable Shape Invariant for Big Data Problems*, *GigaScience*, Volume 12, Article giad094, 2023. Available at: <https://doi.org/10.1093/gigascience/giad094>.
- [4] Paweł Dłotko, Niklas Hellmer, Łukasz Stettner, Rafał Topolnicki, *Topology-Driven Goodness-of-Fit Tests in Arbitrary Dimensions*, *Statistics and Computing*, Volume 34, Article No. 34, 2024.
- [5] Gurjeet Singh, Facundo Mémoli, Gunnar Carlsson, *Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition*, *Eurographics Symposium on Point-Based Graphics*, 2007.
- [6] Paweł Dłotko, Davide Gurnari, Radmila Sazdanovic, *Mapper-Type Algorithms for Complex Data and Relations*, *Journal of Computational and Graphical Statistics*, 2024.
- [7] Paweł Dłotko, Davide Gurnari, Mathis Hallier, Anna Jurek-Loughrey, *ClusterGraph: A New Tool for Visualization and Compression of Multidimensional Data*, *arXiv*, DOI: 10.48550/arXiv.2411.05443, 2024.
- [8] Paweł Dłotko, *Visualizing Maps in Practice*, *Academia*, Images, 2022.
- [9] <https://github.com/dioscuri-tda>

Streszczenia

Zastosowanie modelu AMMI do oceny odmian soi w północnej Polsce

**Konrad Banaś¹, Agnieszka Osiecka², Tomasz Lenartowicz²,
Agnieszka Łącka¹, Henryk Bujak^{2,3}, Marcin Przystalski²**

¹ Uniwersytet Przyrodniczy w Poznaniu

² Centralny Ośrodek Badania Odmian Roślin Uprawnych

³ Uniwersytet Przyrodniczy we Wrocławiu

W doświadczalnictwie rolniczym kluczową kwestią są serie doświadczeń odmianowych, stanowiące szczególny przypadek doświadczeń dwuczynnikowych, w których jednym czynnikiem są odmiany, a drugim miejscowości. Do opisu charakteru interakcji dwóch czynników może być zastosowany wielowymiarowy model statystyczny AMMI, opisujący sumę efektów addytywnych (głównych) odmian i środowisk oraz efekty interakcji GE w multiplikatywnej postaci. Analizie zostały poddane średnie odmianowe trzynastu odmian soi. Model AMMI κ został dopasowany za pomocą pakietu statgenGxE w programie statystycznym R. Zostały wyznaczone średnie oraz wartości indeksu WAAS (weighted averages of the absolute scores) w celu oszacowania stabilności odmian. Ponadto, zostały wyznaczone megaśrodowiska oraz została wykonana analiza adaptacji.

Literatura

- [1] Caliński T., Czajka S., Kaczmarek Z., Krajewski P., Pilarczyk W., Siatkowski I., Siatkowski M., *On mixed model analysis of multi-environment variety trials: a reconsideration of the one-stage and the two-stage models and analyses.*, Statistical Papers, 58 (2017), 433-465.
- [2] Damesa T. M., Møhring J., Worku M., Piepho H. P., *One step at a time: stage-wise analysis of a series of experiments.*, Agronomy Journal, 109 (2017), 845-857.
- [3] Gauch H. G., *Model selection and validation for yield trials with interaction*, Biometrics, 44 (1988), 705-715.
- [4] Gauch H. G., *Statistical analysis of regional yield trials: AMMI analysis of factorial designs.*, Elsevier, Amsterdam, 2009.
- [5] Piepho H. P., Møhring J., Schulz-Streeck T., Ogutu J. O., *A stage-wise approach for analysis of multi-environment trials.*, Biometrical Journal, 54 (2012), 844-860.

Charakteryzacje rozkładów prawdopodobieństwa przez regresje uogólnionych statystyk porządkowych

Mariusz Bieniek

Uniwersytet Marii Curie-Skłodowskiej, Lublin

Rozważamy problem charakteryzacji rozkładów prawdopodobieństwa przez funkcję regresji w modelach uporządkowanych danych statystycznych, takich jak np. statystyki porządkowe, wartości rekordowe itp. Dla ustalonej funkcji ciągłej i ściśle rosnącej $h : (\alpha, \beta) \rightarrow \mathbb{R}$ i dla uogólnionych statystyk porządkowych z ciągłego rozkładu F określamy funkcję regresji

$$\xi(x) = \mathbb{E} \left(h(X_*^{(r+\ell)}) \mid X_*^{(r)} = x \right), \quad x \in (\alpha, \beta),$$

gdzie $r, \ell \geq 1$. Pokażemy, że znajomość funkcji ξ wyznacza rozkład F jednoznacznie i wyznaczymy warunki konieczne i dostateczne na to, aby dana funkcja $\xi : (\alpha, \beta) \rightarrow \mathbb{R}$ była regresją powyższej postaci dla pewnego rozkładu F . Rozwiązanie w przypadku $\ell = 1$ jest znane, w referacie rozważymy przypadek $\ell = 2$.

Estymacja dużych macierzy precyzji i identyfikacja modeli grafowych poprzez regularyzację macierzy korelacji warunkowych

Małgorzata Bogdan¹, Adam Chojecki², Ivan Hejny³,
Bartosz Kołodziejek², Jonas Wallin³

¹ Uniwersytet Wrocławski

² Politechnika Warszawska

³ Lund University

Modele grafowe służą do graficznej reprezentacji struktury zależności między wieloma zmiennymi. W klasycznej reprezentacji modelu grafowego, krawędzie łączą te zmienne, które są warunkowo zależne, pod warunkiem wszystkich pozostałych zmiennych. Oprócz identyfikacji sieci powiązań między zmiennymi, modele grafowe służą również do redukcji wymiaru i zmniejszenia błędu estymacji wysoko-wymiarowych macierzy precyzji i kowariancji. W czasie wykładu zaprezentujemy nową klasę algorytmów do identyfikacji modeli grafowych, opartą na regularyzacji macierzy częściowych korelacji. Podstawową zaletą metod z tej klasy jest naturalna niezależność identyfikacji struktury od skalowania poszczególnych zmiennych. Rozważymy szeroką klasę technik regularyzacyjnych, w tym Partial Correlation Graphical LASSO (PCGLASSO) czy Partial Correlation Graphical SLOPE (PSGSLOPE), które umożliwiają identyfikację nisko-wymiarowego wzorca ograniczającego przestrzeń parametrów macierzy precyzji. Zaprezentujemy szereg wyników teoretycznych, w tym wyniki dotyczące asymptotycznego rozkładu uzyskanych estymatorów w sytuacji, gdy liczba zmiennych jest ustalona, rozmiar próby dąży do nieskończoności, kara maleje w tempie $1/\sqrt{n}$, a rozkłady badanych zmiennych pochodzą z dowolnego rozkładu o eliptycznych konturach i skończonej macierzy kowariancji. Wyniki te umożliwiają wyprowadzenie warunku nie-reprezentowalności (Irrepresentability condition, IR), który gwarantuje asymptotyczną identyfikację wzorca w przypadku nieco silniejszej kary. W szczególności wyniki te pokazują, że w przeciwieństwie do grafowego LASSO (GLASSO), które regularyzuje macierz precyzji, PCGLASSO może skutecznie rozpoznawać grafy generowane przez jedną centralną zmienną (hubs). To zjawisko zostało potwierdzone analizą danych dotyczących ekspresji genów, gdzie PCGLASSO zwraca znacznie bardziej uporządkowaną strukturę niż GLASSO oraz systematycznie wyższą wartość optymalizowanej funkcji wiarygodności.

Szacowanie poziomów zanieczyszczeń środowiskowych

**Jacek Bojarski, Dawid Drapiński, Maciej Niedziela,
Katarzyna Pypka, Jarosław Stein, Gabriel Zaborniak,
Piotr Ziembicki, Kornel Żuber**

Uniwersytet Zielonogórski

Monitorowanie poziomu zanieczyszczeń atmosferycznych jest kluczowym elementem ochrony zdrowia publicznego i środowiska. Zanieczyszczenia takie jak tlenek węgla (CO), dwutlenek azotu (NO₂), dwutlenek siarki (SO₂), ozon (O₃) oraz pyły zawieszone (PM10 i PM2.5) mają znaczący wpływ na jakość powietrza, którym oddychamy. Precyzyjne pomiary tych substancji umożliwiają identyfikację obszarów o wysokim stężeniu zanieczyszczeń, co jest niezbędne dla podejmowania odpowiednich działań zapobiegawczych i redukcyjnych. Pomiary punktowe dostarczają precyzyjnych danych na temat stężenia zanieczyszczeń w konkretnych lokalizacjach, co pozwala na monitorowanie jakości powietrza w wybranych punktach. Niemniej jednak, mieszkańcy są często zainteresowani nie tylko danymi z pojedynczych lokalizacji, ale także ogólnym poziomem zanieczyszczeń, który może występować na większym obszarze, obejmującym ich miejsca zamieszkania, pracy czy codziennych aktywności.

Aby odpowiedzieć na te potrzeby, niezbędne jest modelowanie matematyczne, które pozwala na prognozowanie rozkładu zanieczyszczeń na szerszym obszarze, uwzględniając czynniki takie jak wiatr, topografia terenu, czy różnorodność źródeł emisji. Takie podejście umożliwia tworzenie map zanieczyszczeń, które dostarczają informacji o przewidywanych poziomach zanieczyszczeń w miejscach, gdzie bezpośrednio pomiary nie są dostępne.

W referacie omówione będą elementy szacowania poziomów zanieczyszczeń wykorzystane przez nasz zespół w realizacji projektu pn. „Innowacyjny system monitorowania i analizy jakości powietrza w mieście Żary”.

Bayesian Model Selection for Colored Probabilistic Graphical Models

Adam Przemysław Chojecki¹, Piort Graczyk², Hideyuki Ishi³,
Bartosz Kołodziejek¹

¹ Warsaw University of Technology, Poland

² Université d'Angers, France

³ Osaka City University, Japan

Probabilistic Graphical Models (PGMs) provide a structured framework for representing complex dependencies between random variables. By encoding probabilistic relationships in a graph, where nodes represent variables and edges represent dependencies, PGMs allow the efficient modeling of high-dimensional data.

The lack of an edge between two nodes in a PGM indicates conditional independence, meaning that the two variables are independent of each other, conditioned on the remaining variables. This property enhances interpretability and simplifies computations, as it allows for splitting the joint distribution into smaller, more manageable components.

Colored PGMs (CPGMs) extend this concept by allowing nodes of the same color to share identical distributions. This means that variables with the same color are treated as having the same probabilistic properties, enabling the model to capture symmetries or repeated patterns in the data. By incorporating this structure, CPGMs provide a more flexible and efficient way to model complex systems with variables that are not only conditionally independent but also belong to groups with the same distribution.

"Model Selection" involves choosing the correct graph structure that represents the underlying data-generating process from the data. It is a crucial task that one can interpret as an automatic way of gaining expert knowledge about the considered phenomenon.

Modeling with CPGMs is especially useful in domains such as bioinformatics, social network analysis, and econometrics, where understanding and capturing intricate relationships among multiple variables is crucial.

The Bayesian Model Selection for Gaussian CPGMs on the full graph was developed in [3] and implemented in R as a *gips* package [1], [2]. Current work focuses on extending this idea for a bigger class of graphs.

Literatura

- [1] Chojecki, A.P. and Morgen, P. and Kołodziejek, B., *gips: Gaussian Model Invariant by Permutation Symmetry*, CRAN, 2022, <https://CRAN.R-project.org/package=gips>.
- [2] Chojecki, A.P. and Morgen, P. and Kołodziejek, B., *Learning permutation symmetries with gips in R*, arXiv, 2023, <https://arxiv.org/abs/2307.00790>.
- [3] Graczyk, Piotr and Ishi, Hideyuki and Kołodziejek, Bartosz and Massam, H el ene, *Model Selection in the Space of Gaussian Models Invariant by Symmetry*, The Annals of Statistics, 2022.

Wykrywanie struktur zależności dla pary zmiennych losowych

Bogdan Ćmiel¹, Teresa Ledwina²

¹ Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie

² Instytut Matematyczny Polskiej Akademii Nauk we Wrocławiu

Zajmiemy się problemem identyfikacji zależności między dwiema zmiennymi losowymi. Wprowadzamy nowy estymator funkcji kwantylowej zależności i odpowiednich lokalnych obszarów akceptacji. Prowadzi to do wnikliwej wizualizacji i oceny podstawowej struktury zależności. Proponujemy również nowy test niezależności dwóch zmiennych losowych oparty na wygładzonych estymatorach normy supremum. Zarówno estymatory jak i testy są rangowe, czyli są niezmiennicze względem ściśle rosnących transformacji danych.

Literatura

- [1] Ćmiel B., Ledwina T., *Validation of association*, Insurance: Mathematics and Economics, 91 (2020), 55–67.
- [2] Ćmiel B., Ledwina T., *Detecting dependence structure: visualization and inference*, arXiv:2410.05858, (2024).

Discrete parametric graphical models

Iza Danielewska¹, Bartosz Kołodziejek¹, Jacek Wesołowski¹
Xiaolin Zeng²

¹ Warsaw University of Technology

² Université de Strasbourg

Probability distributions defined on graphical structures offer attractive possibilities for analyzing data with structural dependencies. In this talk, I will introduce four novel classes of graphical distributions for multivariate count data: the multinomial, negative multinomial, multivariate hypergeometric, and multivariate negative hypergeometric distributions. These distributions are defined by graph-multinomial coefficients, which are explicit when the underlying graph is decomposable. I will offer interpretations of these distributions and define their conjugate priors for each model, examining their hyper Markov properties (in the sense of Dawid and Lauritzen [2]). These distributions can be used analogously to their classical counterparts, serving as tools for modeling data with specific dependency structures and opening new perspectives in the statistical analysis of structured count data. The talk is based on joint work with Bartosz Kołodziejek, Jacek Wesołowski, and Xiaolin Zeng.

Literatura

- [1] Lauritzen S.L., *Graphical Models*, Oxford Statistical Science Series, 17, Oxford Science Publications, 1996.
- [2] Dawid A.P., Lauritzen S.L., *Hyper-Markov laws in the statistical analysis of decomposable graphical models*, The Annals of Statistics, 1993, Vol. 21, No. 3, 1272-1317.

Poprawki do praw Zipfa i Heapsa oparte na modelowaniu frakcji hapaksów

Łukasz Dębowski

Instytut Podstaw Informatyki PAN

Prawo Zipfa to najsłynniejsze prawo językoznawstwa ilościowego (lingwistyki kwantytatywnej). Głosi ono, że n -te co do częstości słowo w tekście pojawia się około n razy rzadziej niż słowo najczęstsze [5, 2, 13]. Prawo Heapsa [10, 11, 8, 7] to z kolei pokrewne prawo empiryczne, które głosi, że liczba różnych słów w tekście rośnie w przybliżeniu jak potęga długości tekstu. Ściśle rzecz biorąc, ani prawo Zipfa, ani prawo Heapsa nie modelują dobrze empirycznego brzegowego rozkładu częstości słów. Odstępstwa empirycznego brzegowego rozkładu częstości słów od praw Zipfa i Heapsa są systematyczne i dają się opisać przez dobrze znany nieparametryczny model urnowy [9, 1]. Mianowicie rozkład brzegowy słów w tekście wygląda, jakby słowa były losowane na ślepo bez zwracania z pewnej urny ze słowami. W swoim wystąpieniu przedstawię stosunkowo proste a zarazem dość dokładne parametryczne modele brzegowego rozkładu częstości słów w tekstach dowolnej długości. Pomysł oparty jest na modelowaniu frakcji hapaksów, czyli udziału słów, które pojawiają się jeden raz [3, 6]. Rozkład brzegowy uzyskujemy przez całkowanie i różniczkowanie tej funkcji. Wystąpienie oparte jest na manuskrypcie [4].

Literatura

- [1] R. H. Baayen (2001) *Word frequency distributions*, Kluwer Academic Publishers.
- [2] E. U. Condon (1928) *Statistics of vocabulary*, *Science* 67(1733):300.
- [3] V. Davis (2018). *Types, Tokens, and Hapaxes: A New Heap's Law*, *Glottology* 9(2):113–129.
- [4] Ł. Dębowski (2023) *Corrections of Zipf's and Heaps' Laws Derived from Hapax Rate Models*. <https://arxiv.org/abs/2307.12896>.
- [5] J. B. Estoup (1916) *Gammes sténographiques*, Institut Stenographique de France.
- [6] F. Fengxiang (2010) *An Asymptotic Model for the English Hapax/Vocabulary Ratio*, *Computational Linguistics* 36(4):631–637.
- [7] H. S. Heaps (1978) *Information Retrieval—Computational and Theoretical Aspects*, Academic Press.
- [8] G. Herdan (1964) *Quantitative Linguistics*, Butterworths.
- [9] E. Khmaladze (1988) *The statistical analysis of large number of rare events*. Technical Report MS-R8804. Centrum voor Wiskunde en Informatica, Amsterdam.
- [10] W. Kuraszkiewicz, J. Łukaszewicz (1951) *Ilość różnych wyrazów w zależności od długości tekstu*, *Pamiętnik Literacki* 42(1):168–182.
- [11] B. Mandelbrot (1954) *Structure formelle des textes et communication*, *Word* 10:1–27.
- [12] J. Milička (2009) *Type-token & Hapax-token Relation: A Combinatorial Model*, *Glottology* 2(1):99–110.
- [13] G. K. Zipf (1935) *The Psycho-Biology of Language: An Introduction to Dynamic Philology*, Houghton Mifflin.

Optymalny wybór długości bloku bootstrapowego dla procesów okresowo skorelowanych

Patrice Bertail¹, Anna Dudek²

¹ MODALX, UPL, Univ. Paris Nanterre, CNRS, F92000 Nanterre, Francja

² Akademia Górniczo-Hutnicza w Krakowie, Wydział Matematyki Stosowanej

Omówimy problem wyboru optymalnej długości bloku w przypadku okresowo skorelowanych procesów dla dwóch blokowych metod bootstrap: Uogólnionej Metody Bloków Sezonowych i Rozszerzonej Metody Bloków Ruchomych. Rozważamy dwa problemy estymacji: wartości oczekiwanej oraz średnie sezonowe. W obu przypadkach optymalna długość bloku jest uzyskiwana poprzez minimalizację błędu średnio-kwadratowego odpowiedniego estymatora wariancji bootstrapowej i we wszystkich przypadkach jest ona proporcjonalna do pierwiastka trzeciego stopnia z długości próby. Na koniec przedstawimy wyniki przeprowadzonych symulacji, w których obliczono optymalne długości bloków dla kilku okresowo skorelowanych szeregów czasowych oraz rzeczywiste prawdopodobieństwa pokrycia bootstrapowych przedziałów ufności dla rozważanych parametrów.

Literatura

- [1] Bertail P., Dudek A.E., *Optimal choice of bootstrap block length for periodically correlated time series*, Bernoulli, 30(3) (2024), 2521-2545.

Warunkowe estymatory największej wiarygodności w modelu wielowymiarowym

Katarzyna Filipiak

Politechnika Poznańska

Celem pracy jest wyznaczenie estymatorów największej wiarygodności nieznanymi parametrami w wielowymiarowym modelu liniowym z ograniczeniami nałożonymi na macierz efektów głównych. Pokażemy, że dla różnych typów ograniczeń często problem estymacji może zostać rozwiązany z wykorzystaniem technik optymalizacji wypukłej, prowadząc do dwóch różnych estymatorów, ściśle zależących od nałożonego ograniczenia.

Dla standardowego modelu jednowymiarowego, jako przypadku szczególnego, wskazane zostaną pewne podobieństwa między zaproponowaną metodą estymacji a regresją brzegową. Mimo podobnie sformułowanego problemu pokażemy za pomocą badań symulacyjnych, że proponowane w pracy estymatory mają zwykle lepsze własności niż estymatory wyznaczone z wykorzystaniem regresji brzegowej.

Literatura

- [1] Filipiak K., von Rosen D., Singull M., Rejchel W., *The safety belt estimator under multivariate linear models with inequality constraints*, Submitted.

Graficzny model współwystępowania chorób alergicznych

Konrad Furmańczyk^{1,3}, Wojciech Niemirowicz², Mariola Chrzanowska^{1,3},
Marta Zalewska³

¹ Szkoła Główna Gospodarstwa Wiejskiego w Warszawie

² Uniwersytet Warszawski

³ Warszawski Uniwersytet Medyczny

W referacie zostanie zaprezentowany model graficzny opisujący współwystępowanie chorób alergicznych (por. [1]). Model ten prezentujemy w dwóch wersjach: model generatywny, który odzwierciedla związki przyczynowe oraz przybliżenie modelu generatywnego za pomocą źle wyspecyfikowanego modelu, który jest wydajniejszy obliczeniowo i łatwy do interpretacji. W obu wersjach naszego modelu bierzemy pod uwagę typowe objawy chorób alergicznych i dodatkowe zmienne towarzyszące. Rozważamy dwa skierowane grafy acykliczne (DAG). Pierwszy opisuje informacje o współwystępowaniu chorób alergicznych (zmienne binarne), drugi graf opisuje zależności pomiędzy poszczególnymi objawami a występowaniem tych chorób. W modelu generatywnym krawędzie prowadzą od chorób do objawów, odpowiadających związkom przyczynowym. W źle wyspecyfikowanym modelu odwracamy kierunek krawędzi, które prowadzą od objawów do chorób. Zaproponowany model poddano ocenie korzystając z danych z epidemiologicznego badania dotyczącego występowania chorób alergicznych w Polsce (por. [2]). Do oceny stabilności zaproponowanego modelu użyto techniki bootstrap i jackknife. Nasze wyniki pokazują, że źle wyspecyfikowany model jest dobrym przybliżeniem modelu generatywnego i pomaga przewidzieć częstość występowania chorób alergicznych.

Literatura

- [1] Furmańczyk K., Niemirowicz W., Chrzanowska M., Zalewska M., *Network Model with Application to Allergy Diseases*, Computational Science – ICCS 2024., Proceedings, Part IV, Lecture Notes In Computer Science, 14835 (2024), 105-112.
- [2] Samoliński B., Raciborski F., Lipiec A. et al., *Epidemiologia Chorób Alergicznych w Polsce (ECAP)*,. Alergol Pol., 1(1) (2014), 10-18.

Utracona pojemność ważonych systemów k -spośród- n z dyskretnymi czasami życia komponentów

Agnieszka Goroncy¹, Krzysztof Jasiński¹, Faustyna Korejwo¹,
Marta Rudzate²

¹ Uniwersytet Mikołaja Kopernika w Toruniu, Polska

² Latvian University, Riga, Latvia

Rozważmy ważony system k -spośród- n , w którym każdemu z n komponentów przypisana jest dodatnia liczba całkowita, tzw. waga, interpretowana jako jego całkowita pojemność. System ten jest w stanie sprawności, jeżeli suma wag wszystkich działających komponentów wynosi co najmniej k . Rozważamy przypadek, gdy czasy życia komponentów są modelowane za pomocą dyskretnych zmiennych losowych, które mogą być zależne i pochodzić z różnych rozkładów prawdopodobieństwa. Głównym celem jest wyznaczenie dyskretnego rozkładu zmiennej losowej odpowiadającej pojemności utraconej przez system w momencie jego awarii. Wyprowadzamy dla niej funkcję masy prawdopodobieństwa. Wielkość ta umożliwia m.in. wyznaczanie optymalnej strategii utrzymania sprawności systemu. Podajemy również przykłady numeryczne, ilustrujące wyniki teoretyczne. Powyższe wyniki zostały opisane w pracy [2].

Literatura

- [1] Eryilmaz S., Reliability and performance evaluation of weighted k -out-of- n : G system consisting of components with discrete lifetimes. *Reliab Eng Syst Saf*, 252:110484, 2024.
- [2] Goroncy A., Jasiński K., Korejwo F., Rudzate M., Lost capacity of the weighted k -out-of- n system with discrete component lifetimes. W recenzji, 2024.
- [3] Samaniego F.J., Shaked M., Systems with weighted components. *Stat Probab Lett*, 78:815-823, 2008.
- [4] Wu J.S., Chen R.J., An algorithm for computing the reliability of weighted- k -out-of- n systems. *IEEE Trans Reliab* 43:327-328, 1994.

Estymacja przedziałowa efektu zabiegowego na podstawie danych nieostrych

Przemysław Grzegorzewski

Politechnika Warszawska
Instytut Badań Systemowych PAN

Pojęcie efektu zabiegowego odnosi się do różnicy, która może być przypisana konkretnemu działaniu, odniesionemu do ustalonych warunków kontrolnych. Estymacja efektu zabiegowego jest często celem badań klinicznych czy też innych badań eksperymentalnych w obszarze nauk społecznych, ekonomicznych itp.

W ostatnim czasie rosnącym zainteresowaniem cieszą się metody wnioskowania na podstawie danych przedziałowych. Za pomocą przedziałów można bowiem w prosty sposób modelować brak precyzji, niepewność wynikającą z braku informacji, fluktuacje mierzonej wielkości itp. Niestety, przyjmowana za przestrzeń próbek rodzina przedziałów domkniętych, wyposażona w działania Minkowskiego, nie jest przestrzenią liniową. Sprawia to nie tylko kłopoty rachunkowe, ale utrudnia także konstrukcję narzędzi statystycznych, które byłyby w stanie sprostać specyfice danych przedziałowych. Źródłem dodatkowych trudności jest natura przedziałów losowych, zwłaszcza w ujęciu ontycznym, która niejednokrotnie uniemożliwia stosowanie metod parametrycznych. Stąd też wynika konieczność stworzenia nowych nieparametrycznych metod i technik wspomagających wnioskowanie statystyczne na podstawie danych przedziałowych

W referacie zostanie przedstawiona metoda konstrukcji przedziałów ufności dla efektu zabiegowego na podstawie obserwacji nieostrych, wykorzystująca pewne techniki znane z teorii testów permutacyjnych (por. [1, 2, 3]) oraz niestandardowe działania określone na przestrzeni przedziałów domkniętych (por. [4]), przyjętych jako model nieostrych danych.

Literatura

- [1] Ernst M.D., *Permutation methods: a basis for exact inference*, Statistical Science 19 (2004), 676–685.
- [2] Garthwaite P.H. *Confidence intervals from randomization tests*, Biometrics 52 (1996), 1387–1393.
- [3] Grzegorzewski P., *Permutation k -sample goodness-of-fit test for fuzzy data*, w: Proceedings of the 2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pp. 1-8, 2020.
- [4] Stefanini L., *A generalization of Hukuhara difference and division for interval and fuzzy arithmetic*, Fuzzy Sets and Systems 161 (2010) 1564–1584.

Grafowe modele oddziaływań wyższych rzędów

Aleksandra Grzelik¹, Mateusz Iskrzyński¹, Aleksandra Puchalska²

¹ Instytut Badań Systemowych PAN

² Wydział Matematyki, Informatyki i Mechaniki Uniwersytetu Warszawskiego

Grafy stanowią naturalną reprezentację bezpośrednich relacji pomiędzy obiektami. Hipergrafy pozwalają opisać zależności między grupami trzech lub większej liczby wierzchołków. Joslyn i Nowak (2017) zaproponowali uogólnienie tych struktur nazwane ubergrafem. Pozwala ono na wygodne modelowanie relacji obejmujących zbiory wierzchołków i krawędzi. Ponadto, w ubergrafie, w przeciwieństwie do hipergrafu, możliwe jest przedstawienie wewnętrznej struktury takiego zbioru. Przykładowo, w grafie opisującym sieć reakcji chemicznych, inhibitory i katalizatory osłabiają lub wzmacniają procesy zachodzące pomiędzy innymi związkami. Wprowadzają zatem zależność wyższego rzędu, która jest bardziej adekwatnie reprezentowana poprzez ubergraf niż hipergraf.

W wystąpieniu opowiem o zaletach modelowania opartego na strukturze skierowanego ubergrafu (tzw. pangrafu) oraz o wybranych statystykach tejże struktury. Nacisk zostanie położony na uogólnienie miary centralności Katza. Motywacją do badań nad strukturą pangrafu są ekosystemy charakteryzujące się różnorodnością oraz złożonością oddziaływań międzygatunkowych. W wystąpieniu zaprezentuję empiryczny przykład ekologicznej plantacji kawy, dla którego wykonana została analiza numeryczna.

Literatura

- [1] Cliff Joslyn, Kathleen E. Nowak, *Ubergraphs: A Definition of a Recursive Hypergraph Structure*, ArXiv, 2017.
- [2] Antonio J. Golubski, Erik E. Westlund, John Vandermeer, Mercedes Pascual, *Ecological Networks over the Edge: Hypergraph Trait-Mediated Indirect Interaction (TMII) Structure*, Trends in Ecology & Evolution, 2017

Regularyzacja w regresji różnicowej i modelowaniu przyczynowym

Szymon Jaroszewicz^{1,2}, Krzysztof Rudaś^{1,2}

¹ Instytut Podstaw Informatyki PAN

² Politechnika Warszawska

W pracy zajmujemy się problemem regularyzacji w liniowych modelach regresji różnicowej (*uplift modeling*), służącej do szacowania heterogenicznych efektów danego działania. Rozważamy modele z interakcjami, które są powszechnie stosowane przez statystyków w medycynie i naukach społecznych do szacowania przyczynowego efektu danego działania [1, 3, 6]. Wprowadzamy też nowy typ takiego modelu. Wykazujemy równoważność wszystkich modeli z interakcjami przy braku regularyzacji, a także fakt, że równoważność nie zachodzi dla modeli regularyzowanych. Interakcje w modelu wprowadzają w rozkładzie a-priori ukryte korelacje między współczynnikami dotyczącymi grupy eksperymentalnej i kontrolnej, co nie zostało wcześniej zauważone. Korelacje te zależą od typu modelu z interakcjami, a interpretując regularyzator jako rozkład a-priori, udało nam się wskazać przypadki, kiedy dany rodzaj interakcji jest najbardziej odpowiedni. Ciekawą cechą zaproponowanego nowego typu modelu jest to, że umożliwia on płynne przejście między dwoma typami modeli regresji różnicowej: modelem podwójnym (*double model*) i modelem opartym o przekształcenie zmiennej celu (*transformed target model*) [5]. Nasze wyniki pozostają w mocy zarówno dla regularyzacji typu ridge, jak i Lasso. Eksperymenty na danych syntetycznych w pełni potwierdzają nasze analizy. Porównujemy również przydatność różnych schematów regularyzacji na danych rzeczywistych.

Literatura

- [1] Imai, K., Ratkovic, M.: The annals of applied statistics. Estimating treatment effect heterogeneity in randomized program evaluation **7**, 443–470 (2013)
- [2] Jaśkowski, M., Jaroszewicz, S.: Uplift modeling for clinical trial data. In: ICML 2012 Workshop on Machine Learning for Clinical Data Analysis. Edinburgh (Jun 2012)
- [3] Imbens, G., Rubin, D.: Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction. Cambridge University Press, New York, NY, USA (2015)
- [4] Hernán, M., Robins, J.: Causal Inference. Boca Raton: Chapman & Hall/CRC (2018), forthcoming
- [5] Rudas, K., Jaroszewicz, S.: Linear regression for uplift modeling. Data Mining and Knowledge Discovery **32**(5), 1275–1305 (2018)
- [6] Gross, S.M., Tibshirani, R.: Data shared lasso: A novel tool to discover uplift. Computational Statistics & Data Analysis **101**, 226–235 (2016)

O liczbie komponentów w poszczególnych stanach w trzystanowym systemie k -spośród- n

Krzysztof Jasiński, Agnieszka Goroncy

Uniwersytet Mikołaja Kopernika w Toruniu

Rozważamy trzystanowy system k -spośród- n złożony z komponentów, których czasy życia są niezależnymi, dyskretnymi zmiennymi losowymi o tym samym rozkładzie. Zakładamy, że system podobnie jak jego komponenty może być w trzech stanach: całkowitej awarii, częściowej sprawności, perfekcyjnej działalności. Interesuje nas wektor losowy reprezentujący liczbę komponentów w poszczególnych stanach. Podamy jego dokładny rozkład prawdopodobieństwa. Wynik teoretyczny zostanie zilustrowany przykładem. Zaprezentowane wyniki pochodzą z pracy [1].

Literatura

- [1] Goroncy A., Jasiński K., *Discrete time three-state k -out-of- n system's failure and numbers of components in each state*, Journal of Computational and Applied Mathematics, 457 (2025).

Estymacja parametrów procesu Poissona z ograniczoną funkcją wartości średniej

Alicja Jokiel-Rokita, Ryszard Magiera

Politechnika Wrocawska

Niejednorodne procesy Poissona z ograniczoną funkcją wartości średniej, zwane w literaturze procesami Poissona typu I-go (zob. [1]), są często wykorzystywane w modelowaniu momentów powtarzających się zdarzeń, np. awarii systemów, charakteryzujących się rosnącą niezawodnością. W referacie przedstawimy znane klasyczne metody estymacji oraz propozycje nowych metod estymacji nieznanymi parametrów tych procesów. Pokażemy, że dla pewnych szczególnych przypadków rozważanych procesów, żadna klasyczna metoda nie gwarantuje istnienia oszacowania z prawdopodobieństwem jeden. Jako rozwiązanie problemu nieistnienia oszacowań w tych przypadkach, proponujemy zastosowanie podejścia bayesowskiego z nieinformującym rozkładem a priori. Zastosowanie podejścia bayesowskiego nie tylko gwarantuje istnienie oszacowania z prawdopodobieństwem jeden, ale również, co wynika z symulacji, prowadzi do bardziej dokładnych, w sensie błędu średniokwadratowego, oszacowań parametrów, w większości rozważanych w symulacjach przypadków.

Literatura

- [1] Kuo L., Yang T. Y. *Bayesian computation for nonhomogeneous Poisson processes in software reliability*, JASA, 91 (1996), 763-773.

Integracja różnych czynników doświadczalnych i grup cech w zaawansowanych modelach sieciowych roślin

Marta Kańczurzevska-Staszak¹, Aneta Sawikowska^{2,3}

¹ Politechnika Poznańska

² Uniwersytet Przyrodniczy w Poznaniu

³ Xenstats sp. z o.o.

Uzyskanie kompleksowego zrozumienia złożonych reakcji biologicznych w badaniach nad roślinami stało się możliwe dzięki integracji różnych poziomów czynników doświadczalnych oraz wielu grup danych multiomicznych (transkryptomicznych, proteomicznych, metabolomicznych, lipidomicznych). Coraz większa liczba poziomów czynników doświadczalnych, typów tkanek oraz punktów czasowych w badaniach sprawia, że zrozumienie wieloczynnikowych zależności staje się bardziej wymagające. Niniejsze badanie koncentruje się na zaawansowanych metodach analizy sieci korelacyjnych oraz na porównaniu tych sieci dla różnych poziomów czynnika doświadczalnego lub dla kombinacji poziomów dwóch bądź więcej czynników, na przykładzie między innymi żyta, jęczmienia i pszenicy.

W ramach analizy zastosowano integrację danych opartą na WGCNA (Ważonej Analizie Współekspresji Genów) w środowisku R, zgodnie z metodami opisanymi przez [1] i [2]. Dodatkowo, wykorzystano sieci korelacyjne różnicowe, zaprezentowane przez [3], oraz modele graficzne Gausse'a za pomocą pakietu `rags2ridges` według [4]. Nowo opracowane metody pozwoliły na analizę złożonych wzorców korelacyjnych w różnych warunkach eksperymentalnych, wychodząc poza standardowe podejście parowe.

Sieci korelacyjne i różnicowe sieci korelacyjne skonstruowano aby zidentyfikować zależności specyficzne dla czynników doświadczalnych i cech, co pozwoliło na wyeksponowanie różnic w powiązaniach między cechami w odmiennych warunkach. Te złożone sieci zwizualizowano w programie Cytoscape, co umożliwiło dokładną i efektywną interpretację wyników. Dodatkowo, modele graficzne Gausse'a pozwoliły na zidentyfikowanie unikalnych połączeń specyficznych dla każdego rodzaju czynnika doświadczalnego oraz tych wspólnych dla wielu warunków, a uzyskane w ten sposób sieci zostały zwizualizowane w R.

Literatura

- [1] Langfelder P., Horvath S. *WGCNA: an R package for weighted correlation network analysis*, BMC Bioinformatics, (9):559 (2008).
- [2] Langfelder P., Zhang B., Horvath S. *Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut library for R*, Bioinformatics, 24(5) (2011), 719–720.
- [3] Piasecka A., Sawikowska A., Kuczyńska A., Ogrodowicz P., Mikołajczak K., Krystowiak K., Gudyś K., Guzy-Wróbelska J., Krajewski P., Kachlicki P. *Drought related secondary metabolites of barley (*Hordeum vulgare* L.) leaves and their association with mQTLs*, Plant Journal, 89 (2017), 898–913.
- [4] Peeters C.F.W., Bilgrau A.E., N. van Wieringen W. *rags2ridges: A One-Stop-2-Shop for Graphical Modeling of High-Dimensional Precision Matrices*. *Journal of Statistical Software*, 102(41) (2022), 1–32.

Pattern recovery with GSLOPE

Bartosz Kołodziejek

Politechnika Warszawska

In my presentation, I will discuss the problem of recovering the pattern of the precision matrix in graphical models within a high-dimensional setting. The notion of pattern depends on the choice of the (non-differentiable) penalty imposed on model's complexity. For the L_1 penalty (LASSO), the pattern consists of the signs (positive, negative, or zero) of the off-diagonal entries of the precision matrix. The problem of high-dimensional covariance estimation and pattern recovery through graphical LASSO was examined in [1].

When using the SLOPE penalty (a generalization of the LASSO penalty), the patterns become much complex and, in addition to signs, also includes clusters (sets of entries of the precision matrix with equal absolute values) and the hierarchy among clusters. Such patterns describe colored graphical models defined in [2]. By employing and enhancing techniques from [1], we establish a sufficient condition for SLOPE pattern recovery, which allows us to determine the minimal sample size n for which pattern consistency is highly probable. By interpreting the Gaussian log-likelihood as a Bregman divergence for log det function, our results apply to any distribution of random vectors with a finite second moment. An important byproduct of our work is the formulation of a quantitative criterion for selecting the sequence of tuning parameters in the SLOPE penalty in an optimal way to maximize the probability of pattern recovery.

This presentation is based on joint work with Piotr Graczyk (Angers), Hideto Nakashima (Tokyo), and Maciej Wilczyński (Wrocław).

Literatura

- [1] Ravikumar, P., Wainwright, M. J., Raskutti, G., Yu, B. *High-dimensional covariance estimation by minimizing ℓ_1 -penalized log-determinant divergence*. Electron. J. Stat. 5 (2011), 935–980
- [2] Hojsgaard, S., Lauritzen, S. L. *Graphical Gaussian models with edge and vertex symmetries*. J. R. Stat. Soc. Ser. B Stat. Methodol. 70 (2008), no. 5, 1005–1027.

Wpływ strategii wstępnego przetwarzania danych na jakość modeli opartych na drzewach w kontekście AutoML

Hubert Ruczyński, Anna Kozak

Politechnika Warszawska

Liczba rozwiązań i algorytmów uczenia maszynowego rośnie w szybkim tempie każdego roku. Ze względu na nadmiar informacji, naukowcy zajmujący się danymi coraz bardziej ufają własnemu doświadczeniu lub powszechnej wiedzy. Celem naszego badania jest ocena, które strategie wstępnego przetwarzania danych najlepiej wpływają na jakość modelu. W swojej pracy skupiamy się na modelach opartych na drzewach [1], badanie obejmuje analizę 38 różnych strategii wstępnego przetwarzania danych i wytrenowanie prawie miliona modeli opartych na drzewach [2], [3]. Analizujemy wpływ różnych strategii przygotowania danych i przedstawiamy te najlepiej działające z wykorzystaniem nowo wprowadzonej miary wstępnego przetwarzania.

Literatura

- [1] Grinsztajn L., Oyallon E., Varoquaux G., *Why do tree-based models still outperform deep learning on typical tabular data?* In Proceedings of the 36th International Conference on Neural Information Processing Systems (NIPS '22), 2022.
- [2] Ruczyński H., Kozak A., *Do Tree-based Models Need Data Preprocessing?*, AutoML Conference 2024 (Workshop Papers), 2024.
- [3] Kozak A., Ruczyński H., *forester: A Novel Approach to Accessible and Interpretable AutoML for Tree-Based Modeling*, AutoML Conference 2023 (Workshop Papers), 2023.

Szeregi czasowe stacjonarne o własnościach cyklicznych

Łukasz Lenart

Uniwersytet Ekonomiczny w Krakowie

W pracach [2] oraz [3] wykorzystano konstrukcję cyklicznych procesów stacjonarnych $\{y_t\}_{t \in \mathbb{Z}}$ zakładającą (dla uproszczenia rozważmy tylko jedną częstotliwość $\omega \in (0, \pi]$):

$$y_t = \alpha_t \cos \omega t + \beta_t \sin \omega t, \quad (1)$$

gdzie $\{\alpha_t\}_{t \in \mathbb{Z}}$ oraz $\{\beta_t\}_{t \in \mathbb{Z}}$ to procesy Gaussowskie, nieskorelowane, stacjonarne, o zerowej wartości oczekiwanej i o takiej samej funkcji autokowariancji $\{\gamma(\tau)\}_{\tau \in \mathbb{Z}}$. Gwarantuje to stacjonarność procesu $\{y_t\}_{t \in \mathbb{Z}}$ o pseudo-cyklicznej funkcji autokowariancji $E(y_t y_{t+\tau}) = \gamma(\tau) \cos(\tau\omega)$. Przy tych założeniach wartość oczekiwana procesu amplitudy $a_t = \sqrt{\alpha_t^2 + \beta_t^2}$ jest proporcjonalna do odchylenia standardowego tego procesu (patrz szersza dyskusja w [1]), co nie jest akceptowalne dla większości zastosowań empirycznych. Prawdopodobnie przez to koncepcja ta nie znalazła wielu zastosowań.

W pracy przedstawiono szereg własności teoretycznych procesów cyklicznych postaci (1), które nie bazują na Gaussowskości dla $\{\alpha_t\}_{t \in \mathbb{Z}}$ oraz $\{\beta_t\}_{t \in \mathbb{Z}}$, również te z wieloma częstotliwościami. Bazując na alternatywnej reprezentacji

$$y_t = A_t \cos(\omega(t + P_t)),$$

gdzie A_t to proces amplitudy, zaś P_t to proces przesunięcia w fazie, wykazano że przy odpowiednich założeniach proces ten jest procesem stacjonarnym lub prawie okresowo skorelowanym. Wyznaczono postać momentów wszystkich rzędów.

Literatura

- [1] Lenart Ł., *Properties of stationary cyclical processes*, arXiv preprint arXiv:2405.08907, 2024.
- [2] Hannan, E.J., Terrell, R. D., and Tuckwell, N.E., *The seasonal adjustment of economic time series*, *International Economic Review*, (1970), 11(1): 24-52.
- [3] Proietti, T., Maddanu, F., *Modelling cycles in climate series: The fractional sinusoidal waveform process*, *Journal of Econometrics*, 2023, <https://doi.org/10.1016/j.jeconom.2022.04.008>

Analiza spektralna i subsampling dla procesów spektralnie skorelowanych

Anna E. Dudek, Bartosz Majewski

AGH Akademia Górniczo-Hutnicza

Procesy spektralnie skorelowane są procesami harmonizowalnymi, których miara spektralna jest skoncentrowana na policzalnym zbiorze krzywych. W naszych badaniach skupiamy się na analizie spektralnej takich procesów w przypadku, gdy krzywe są prostymi o nachyleniach innych niż jednostkowe. Tego rodzaju procesy mają praktyczne zastosowania, np. w lokalizowaniu ruchomych źródeł, takich jak samoloty, rakiety lub inne wrogi urządzenia zakłócające emitujące sygnały komunikacyjne. Naszym celem jest estymacja funkcji gęstości spektralnej. Jako jej estymator przyjmujemy periodogram wygładzony wzdłuż rozważanej prostej. Podczas prezentacji przedstawimy wyniki dotyczące jego zgodności w sensie średniokwadratowym i rozkładu asymptotycznego. Następnie omówimy asymptotyczne właściwości estymatora koherencji. Ponadto zaproponujemy technikę subsamplingu i pokażemy jej zgodność w rozważanym problemie. W ostatniej części prezentacji przedstawiamy symulacje, które ilustrują uzyskane wyniki teoretyczne.

Literatura

- [1] Dudek, A. E., Majewski, B., *Asymptotic distribution and subsampling in spectral analysis for spectrally correlated processes*, Submitted manuscript, (2024), <https://hal.science/hal-04675084>

Bootstrap for high-dimensional Markov chains

Patrice Bertail¹, Anna Dudek², Karolina Marek²

¹ UFR SEGMI, MODAL'X, Université Paris-Nanterre 200, Paris, France

² AGH University of Krakow, Department of Applied Mathematics

Large-dimensional Markov chains appear in many models, either due to the simultaneous study of multiple time series or because the time series are Markov chains of order $p > 1$, resulting in multidimensional data after proper vectorization. We will introduce the Multidimensional Approximative Regenerative Block Bootstrap (MARBB), a bootstrap algorithm designed for high-dimensional Markov chains. We will focus on a vector autoregressive (VAR(1)) process with a low-rank structure. We propose a reduction algorithm that transforms the original high-dimensional time series into a lower-dimensional Markov chain. Once the chain is reduced, we leverage the regenerative properties of Harris recurrent Markov chains within a general state space, using the Nummelin splitting technique to extend results from the one-dimensional case (see [1]) to the multidimensional setting. This approach enables the identification of the regeneration times of the lower-dimensional Markov chain, which in turn leads to the splitting of the original high-dimensional Markov chain into regenerative blocks. We will present the MARBB consistency results, and we will demonstrate an example of the application of our algorithm using simulation data.

Literatura

- [1] Bertail, P., Cléménçon, S. (2006). *Regeneration-based statistics for Harris recurrent Markov chains*. In: Bertail, P., Soulier, P., Doukhan, P. (eds) *Dependence in Probability and Statistics. Lecture Notes in Statistics*, vol 187. Springer, New York, NY.
- [2] Bertail, P., Dudek, A., Marek, K. *Bootstrap for high-dimensional Markov chains*. Working paper.

Testy statystyczne dla indeksu Jaccarda

Błażej Miasojedow

Uniwersytet Warszawski

Popularnym sposobem opisywania obiektów jest przedstawienie jego cech za pomocą wektorów binarnych. W takiej sytuacji badanie podobieństwa pomiędzy obserwacjami sprowadza się do oceny odległości pomiędzy binarnymi wektorami. Typowym sposobem miary podobieństwa jest indeks Jaccarda.

W swoim referacie opowiem o metodach statystycznych pozwalających na podstawie indeksu Jaccarda ocenić czy w sensie statystycznej istotności dwa wektory binarne są podobne czy nie. Referat będzie oparty na pracy [1].

Literatura

- [1] Chung, N., Miasojedow, B., Startek, M. & Gambin, A. *Jaccard/Tanimoto similarity test and estimation methods for biological presence-absence data*. BMC Bioinformatics. **20**, 644 (2019,12,24), <https://doi.org/10.1186/s12859-019-3118-5>

Weryfikacja założenia SCAR (Selected Completely at Random) dla danych typu PU (Positive Unlabeled)

P. Teisseyre^{2,1}, K. Furmańczyk³, J. Mielniczuk^{2,1}

¹ Wydział Matematyki i Nauk Informatycznych, Politechnika Warszawska

² Instytut Podstaw Informatyki, Polska Akademia Nauk

³ Instytut Informatyki Technicznej, Szkoła Główna Gospodarstwa Wiejskiego

Jednym z celów uczenia się dla danych typu PU (Positive Unabeled) jest konstrukcja klasyfikatora na podstawie danych składających się z obserwacji z klasy pozytywnej i obserwacji nieetykietowanych, które mogą należeć do klasy pozytywnej lub do negatywnej. Modelowanie danych PU wymaga założeń o mechanizmie etykietowania: najprostszym jest założenie SCAR (Selected Completely at Random), które zakłada, że etykietowanie w klasie dodatniej jest niezależne od wartości predyktorów, w odróżnieniu od bardziej ogólnego założenia SAR (Selected at Random), które taką zależność dopuszcza. W pracy [1] proponujemy prosty i obliczeniowo nieobciążający test spełnialności założenia SCAR. Taka weryfikacja ma znaczenie praktyczne: klasyfikatory skonstruowane przy założeniu SCAR są, w przypadku spełnienia tego założenia, bardziej efektywne i zdecydowanie szybsze niż klasyfikatory skonstruowane dla ogólnego modelu PU-SAR.

Literatura

- [1] Furmańczyk, K., Teisseyre, P., Mielniczuk, J., *Verifying the Selected Completely at Random Assumption in Positive-Unlabeled learning*, Proceedings of XXVIIIth European Conference of Artificial Intelligence ECAI'2024, p. 1672-1680.

Likelihood ratio test for covariance matrix under the multivariate and matrixvariate t distribution

Katarzyna Filipiak¹, Daniel Klein², Stepan Mazur³,
Malwina Mrowińska¹

¹ Poznań University of Technology

² P. J. Šafárik University

³ Örebro University

One of the most common distributional assumptions of many statistical models is the normal distribution. However, in many applied fields such as, for example, macro- and financial economics, heavy tailed distributions are more appropriate to consider. In this talk one of the most popular heavy-tailed distributions, the multivariate t distribution, is considered. Additionally, in this talk we relax the assumption about independence to uncorrelated observations, which is indistinguishable under normality assumption. The aim of this talk is to present the maximum likelihood estimators of location and scale parameters of multivariate t distribution, however, since the observations are uncorrelated (not independent), for this purpose the matrix t distribution will be used. Observe, that the definition of matrix t distribution depends on the stochastic representation of matrix-variate t distribution random variable and thus two alternative forms of distribution will be considered.

Moreover, the likelihood ratio test is proposed and its distributional properties are verified under the null hypothesis with fully precised covariance matrix or precised up to the constant.

Porównanie metod modelowania zbiorów szeregów czasowych i oceniania ich skuteczności

Karol Opara

Instytut Badań Systemowych PAN

W refracie przedstawię różne metody modelowania predykcyjnego szeregów czasowych. Motywację i ilustrację moich rozważań stanowią będą dane z badania klinicznego mającego na celu detekcję zmiany fazy u pacjentów cierpiących na chorobę afektywną dwubiegunową za pomocą aplikacji zainstalowanej na ich smartfonach [1].

Zagadnienie predykcyjne sformułować można zarówno jako regresję jak i klasyfikację [2, 3]. Z natury problemu, zmienna objaśniana zawiera znacznie więcej „braków” niż predyktory. Skłania to do rozważenia metod częściowo nadzorowanych [4, 5] albo wręcz nie-nadzorowanych, opartych na analizie skupień bądź też badaniu stabilności procesów [5]. Przedstawię też zalety i wady tworzenia wspólnego modelu dla wszystkich pacjentów oraz częściowo bądź w pełni zindywidualizowanych modeli.

Różnorodność sposobów sformułowania zadania detekcji zmiany fazy utrudnia jednorodną ocenę ich jakości. Występuje też potrzeba blokowej walidacji krzyżowej, która jednak często jest zaniedbywana w literaturze przedmiotu.

Literatura

- [1] Antosik-Wójcińska, A. Z., Dominiak, M., Chojnacka, M., Kaczmarek-Majer, K., Opara, K. R., Radziszewska, W., ... & Świącicki, Ł. (2020). *Smartphone as a monitoring tool for bipolar disorder: a systematic review including data analysis, machine learning algorithms and predictive modelling*. International Journal of Medical Informatics, 138, 104131.
- [2] Dominiak, M., Kaczmarek-Majer, K., Antosik-Wójcińska, A. Z., Opara, K. R., Olwert, A., Radziszewska, W., ... & Mierzejewski, P. (2022). *Behavioral and self-reported data collected from smartphones for the assessment of depressive and manic symptoms in patients with bipolar disorder: prospective observational study*. Journal of Medical Internet Research, 24(1), e28647.
- [3] Kaczmarek-Majer, K., Dominiak, M., Antosik, A. Z., Hryniewicz, O., Kamińska, O., Opara, K., ... & Świącicki, Ł. (2024). *Acoustic features from speech as markers of depressive and manic symptoms in bipolar disorder: A prospective study*. Acta Psychiatrica Scandinavica.
- [4] Kaczmarek-Majer, K., Casalino, G., Castellano, G., Hryniewicz, O., & Dominiak, M. (2022). Explaining smartphone-based acoustic data in bipolar disorder: Semi-supervised fuzzy clustering and relative linguistic summaries. *Information Sciences*, 588, 174-195.
- [5] Casalino, G., Castellano, G., Hryniewicz, O., Leite, D., Opara, K., Radziszewska, W., & Kaczmarek-Majer, K. (2023). *Semi-Supervised vs. Supervised Learning for Mental Health Monitoring: A Case Study on Bipolar Disorder*. International Journal of Applied Mathematics and Computer Science, 33(3), 419-428.

Wpływ struktury systemu na przenoszenie własności starzenia jego komponentów

Jagoda Papis¹, Tomasz Rychlik², Magdalena Szymkowiak¹

¹ Politechnika Poznańska
² Polska Akademia Nauk

Navarro i Rubio [2] przedstawili struktury wszystkich 180 systemów koherentnych złożonych z 5 komponentów. Podstawowe własności starzenia systemów i warunki, które określają czy dana własność jest dziedziczona przez system zostały omówione w pracach [1] i [3].

Obecnie badamy podstawowe własności wszystkich systemów o 5 komponentach o różnej strukturze przy założeniu, że komponenty są niezależne i o jednakowym rozkładzie (I.I.D.). Przegląd tych własności dla poszczególnych systemów pozwoli zbadać wpływ struktury systemu o dowolnej liczbie komponentów na dziedziczenie własności starzenia. Szczególną uwagę poświęcamy własności rosnącej intensywności uszkodzeń (IFR).

Literatura

- [1] Navarro J., *Introduction to System Reliability Theory*, Springer, 2022.
- [2] Navarro J., Rubio R., *Computations of Signatures of Coherent Systems with Five Components*, Communications in Statistics - Simulation and Computation, 39:1 (2009), 68-84.
- [3] Rychlik T., Szymkowiak M., *Signature Conditions for Distributional Properties of System Lifetimes if Coherent Lifetimes are I.I.D. Exponential*, IEEE Transactions on Reliability, 71:2 (2022), 590-602.

Estimation of conditional inequality measures

Alicja Jokiel-Rokita¹, Sylwester Piątek¹, Rafał Topolnicki²

¹ Politechnika Wrocławska

² Polska Akademia Nauk

Classical inequality measures such as the Gini index are often used to describe the sparsity of the distribution of a certain feature in a population. It is sometimes also used to compare the inequalities between some subpopulations, conditioned on certain values of the covariates. The concept of measuring inequality in subpopulation was described in the literature (see, for example, [5]) and it is strongly related to the decomposition of the Gini index [3].

We discuss an idea of conditional inequality measure extended to the case when covariates are continuous. Curves of conditional inequality measures are introduced, especially, the curves of the conditional quantile versions of the Zenga and D indices are considered [4]. Various methods of their estimation based on quantile regression are presented. A novel approach using isotonic regression is used to obtain quantile regression lines which lead to monotonous estimators of the conditional quantile function as an alternative for several known methods, for example proposed by Bassett and Koenker [1] or by Bondell et al. [2]. Asymptotic properties of the new estimators are shown. The accuracy of the estimators considered is compared in simulation studies and an example application is given. The simulation study shows that our new approach outperform the standard methods in most of the cases examined in this study.

The talk is based on a paper, which is currently in preparation.

Literatura

- [1] Bassett G., Koenker R. *An empirical quantile function for linear models with iid errors*, Journal of the American Statistical Association, 77, 378 (1982), 407–415.
- [2] Bondell H. D., Reich B. J., Wang H. *Noncrossing quantile regression curve estimation*, Biometrika, 97, 4 (2010), 825–838.
- [3] Dagum C., *A new approach to the decomposition of the Gini income inequality ratio*, Income Inequality, 22 (1997), 515–531.
- [4] Jokiel-Rokita A., *Nonparametric estimators of inequality curves and inequality measures*, preprint (2023), arxiv.org/abs/2310.16200.
- [5] Larraz B., *Decomposing the Gini Inequality Index: An Expanded Solution With Survey Data Applied to Analyze Gender Income Inequalit*, Sociological Methods & Research, 44(3) (2015), 508–533.

Analiza czasu przeżycia pacjentów z ostrym zespołem wieńcowym

Katarzyna Pypka, Magdalena Wojciech, Konrad Pieszko

Uniwersytet Zielonogórski

Ostry zespół wieńcowy (OZW) stanowi istotne wyzwanie dla współczesnej medycyny, a zrozumienie czynników wpływających na przeżycie pacjentów jest kluczowe dla skutecznego leczenia. Analiza przeżywalności pacjentów opiera się na wynikach badania przeprowadzonego na 1 042 osobach chorych na OZW, które w trybie nagłym zostały przyjęte na oddział kardiologiczny. Zastosowano metody statystyczne, takie jak estymator Kaplana-Meiera i model proporcjonalnych hazardów Coxa.

Podejście modelowe z wykorzystaniem metod analizy przeżycia pozwoliło na wyodrębnienie istotnych czynników klinicznych takich jak markery krwi, wiek oraz płeć wpływających na długość życia chorych.

Uzyskane wyniki mają potencjał do optymalizacji i poprawy strategii terapeutycznych, co w efekcie może wpływać na lepsze dostosowanie leczenia i poprawę rokowań pacjentów.

W prezentacji zaprezentowane zostaną różne możliwości przyszłych badań wspierających diagnostykę pracy serca oparte na analizie przepływu krwi w sercu na podstawie obrazu z tomografu komputerowego.

Prezentacja przedstawia nowe kierunki badań nad diagnostyką pracy serca, wykorzystujące analizę przepływu krwi na obrazach z tomografii komputerowej. Umożliwi to dokładniejsze monitorowanie pracy serca i wykrywanie zagrożeń.

Literatura

- [1] Stevenson M., *An introduction to survival analysis*, EpiCentre, 2007.
- [2] Pieszko K., *Machine-learned models using hematological inflammation markers in the prediction of short-term acute coronary syndrome outcomes*, *Journal of Translational Medicine*, 16(1):334, 2018.
- [3] Otani, T., Al-Issa, A., Pourmorteza, A., McVeigh, E. R., Wada, S., Ashikaga, H. *A Computational Framework for Personalized Blood Flow Analysis in the Human Left Atrium*. *Annals of biomedical engineering*, 44(11), 3284–3294, 2016.

Łączna estymacja prawdopodobieństwa a posteriori i funkcji skłonności w modelu PU

Jan Mielniczuk^{1,2}, Wojciech Rejchel³, Paweł Teisseyre^{1,2}

¹ Polska Akademia Nauk

² Politechnika Warszawska

³ Uniwersytet Mikołaja Kopernika

W klasyfikacji binarnej należy przewidzieć, do której z dwóch klas należy badany obiekt na podstawie obserwacji zmiennych pomocniczych tego obiektu oraz zbioru danych $(X_1, Y_1), \dots, (X_n, Y_n)$, gdzie $X_i \in \mathbb{R}^p$ oraz $Y_i \in \{-1, 1\}$. Jednakże często zdarza się, że zmienne Y_i nie są obserwowane, na przykład w modelu PU (Positive-Unlabeled) zakłada się, że dane są postaci $(X_1, S_1), \dots, (X_n, S_n)$ oraz $S_i \in \{0, 1\}$. Jeśli $S_i = 1$, to $Y_i = 1$, ale $S_i = 0$ oznacza, że obiekt ten nie posiada etykiety, czyli Y_i jest nieznanne. Problem ten dość powszechnie występuje w praktyce, na przykład podczas analizy danych medycznych bądź genetycznych.

Standardowym założeniem wielu algorytmów działających w modelu PU jest restrykcyjny warunek, że funkcja skłonności $e(x) = P(S = 1|Y = 1, X = x)$ jest stała. Chcąc uniknąć tego założenia, rozważamy podejście parametryczne, w którym prawdopodobieństwo a posteriori oraz funkcja skłonności są modelowane jak w regresji logistycznej (model podwójnie logistyczny). Zaproponowany algorytm polega na minimalizacji ryzyka empirycznego, które zależy od obu tych funkcji. W referacie zostaną omówione jego własności teoretyczne i praktyczne.

Zamiana wartości klas w modelowaniu przyczynowości dla niezbalansowanych zrandomizowanych danych

Krzysztof Rudaś^{1,2}, Szymon Jaroszewicz^{1,2}

¹ Politechnika Warszawska

² Instytut Podstaw Informatyki PAN

Modelowanie przyczynowości zajmuje się oszacowaniem efektu przeprowadzonego działania (np. kampanii marketingowej, terapii) na poszczególne jednostki. W tym celu dzielimy populację na grupę eksperymentalną i kontrolną i szacujemy różnicę odpowiedzi pomiędzy tymi dwiema grupami. Brak zbalansowania zmiennej odpowiedzi i zmiennej opisującej przydział do grupy eksperymentalnej i kontrolnej może wpłynąć znacząco na jakość uzyskiwanych przewidywań. Jednym ze sposobów rozwiązania tego problemu są metody oparte na undersamplingu [4]. Niestety tego typu podejścia zniekształcają predykcję efektu działania. W prezentacji przedstawię alternatywną metodę, polegającą na zamianie wartości klas binarnej zmiennej odpowiedzi dla wybranych obserwacji. Metoda ta nie zniekształca wartości predykcji i działa szczególnie dobrze dla przyczynowych modeli bazujących na przekształceniu zmiennej odpowiedzi [2]. Oprócz wyników teoretycznych, przedstawię rezultaty uzyskane dla zrandomizowanych danych rzeczywistych.

Literatura

- [1] Nyberg O., Klami A., *Exploring uplift modeling with high class imbalance*, Data Mining and Knowledge Discovery, Springer, 37 (2023), 736-766.
- [2] Rudaś K., Jaroszewicz, S. *Linear regression for uplift modeling*, Data Mining and Knowledge Discovery, Springer, 32 (2018), 1275–1305.

Autocovariance estimation for generalized almost-cyclostationary processes with unknown lag-dependent cycle frequencies

Anna Dudek¹, Antonio Napolitano², Jakub Rutkowski¹, Jakub Wojdyła¹

¹ Faculty of Applied Mathematics, AGH University of Krakow

² Parthenope University of Naples, Italy

The generalized almost-cyclostationary processes (GACS) is a class of continuous stochastic processes, for which the statistical moments are almost-periodic functions of time, for every fixed delay. For such processes, the autocovariance function can be expressed as a trigonometric polynomial, with functions of delay used as frequencies. These functions are called lag-dependent cycle frequencies. The problem of estimating autocovariance function (or more precisely the generalised cycle autocovariance function), with the unknown lag-dependent cycle frequency, will be addressed. The mean-square consistency of the cyclic correlogram, with estimated lag-dependent cycle frequency (as estimator of the generalized cycle autocovariance function), will be considered. A method for estimating lag-dependent cycle frequencies will be presented. Moreover, the performance of the cyclic correlogram with estimated lag-dependent cycle frequencies will be evaluated through numerical simulation.

Literatura

- [1] Napolitano A., *Cyclostationary Processes and Time Series. Theory, Applications, and Generalizations*, Elsevier, 2019.
- [2] Napolitano A., *Generalizations of Cyclostationary Signal Processing: Spectral Analysis and Applications*, John Wiley & Sons Ltd. – IEEE Press, 2012.

Warunki skończoności i oszacowania momentów uogólnionych statystyk pozycyjnych

Mariusz Bieniek¹, Tomasz Rychlik²

¹ Uniwersytet Marii Curie-Skłodowskiej

² Instytut Matematyczny PAN

Semiparametryczny model uogólnionych statystyk pozycyjnych to ogólny opis rozkładów uporządkowanych zmiennych losowych obejmujący w szczególności klasyczne statystyki pozycyjne, klasyczne i k te wartości rekordowe, obserwacje ze schematu progresywnego cenzurowania niezależnych zmiennych losowych o jednakowym rozkładzie i wiele innych. Elementami modelu są wektor lub ciąg nieujemnych parametrów oraz jednowymiarowa dystrybuanta zwana bazową.

W referacie określamy warunki konieczne i dostateczne skończoności momentów ustalonego rzędu uogólnionych statystyk pozycyjnych z ustalonym wektorem parametrów oraz dowolną bazową dystrybuantą o nieujemnym nośniku i skończoną wartością oczekiwaną. Warunki te zależą od relacji między rzędem momentu a najmniejszym parametrem modelu uogólnionych statystyk pozycyjnych i jego krotnością. Ponadto, przy zachowaniu tych warunków, wyznaczamy dokładne górne oszacowania momentów uogólnionych statystyk pozycyjnych mierzonych w jednostkach skali będących odpowiednimi potęgami wartości oczekiwanej rozkładu bazowego.

Zastosowanie metod wielowymiarowych w badaniu wpływu makuchu lnianego na bazylię w warunkach suszy

Aneta Sawikowska^{1,2}, Marta Kańczurzevska-Staszak³

¹ Uniwersytet Przyrodniczy w Poznaniu

² Xenstats sp. z o.o.

³ Politechnika Poznańska

Omówione zostaną wyniki badań nad wpływem makuchu lnianego na wzrost, skład chemiczny oraz adaptację bazylii (*Ocimum basilicum* L.) w warunkach stresu suszy, ze szczególnym uwzględnieniem metod wielowymiarowych, takich jak analiza korelacji oraz analiza składowych głównych (PCA) na podstawie [1]. Badanie miało na celu sprawdzenie, czy makuch lniany jako ekologiczny dodatek do podłoża może zwiększyć odporność bazylii na niedobory wody, a jednocześnie poprawić jej wydajność i jakość plonów.

W badaniu skonstruowano sieci korelacyjne oraz różnicową sieć korelacyjną, aby przeanalizować zależności między kluczowymi cechami bazylii, takimi jak masa roślin, zawartość związków bioaktywnych i skład chemiczny olejków eterycznych, w różnych warunkach nawadniania. Przy użyciu pakietu WGCNA w R ([2]) macierze korelacji Pearsona zostały przekształcone w macierze sąsiedztwa, co pozwoliło następnie na identyfikację modułów, czyli grup silnie powiązanych zmiennych. Zidentyfikowano także tzw. „huby” — kluczowe zmienne o największej liczbie połączeń z innymi cechami, co wskazuje na ich kluczową rolę w mechanizmach adaptacyjnych rośliny. Różnicowe sieci korelacyjne, z wykorzystaniem transformacji Z Fishera oraz korekcji Bonferroniego [3], pozwoliły na identyfikację istotnych statystycznie zmian w korelacjach między warunkami suszy i nawodnienia. Wszystkie sieci zostały zwizualizowane za pomocą programu Cytoscape.

Analiza PCA została przedstawiona za pomocą biplotu, który umożliwił wizualizację wpływu różnych dawek makuchu lnianego na zmienne chemiczne i fizjologiczne bazylii. Biplot ukazał zależności między grupami zmiennych oraz ich rozkład w przestrzeni wyznaczonej przez dwa główne komponenty. Umożliwiło to wizualne przedstawienie wpływu różnych dawek makuchu na skład olejków eterycznych oraz parametry wzrostowe bazylii w warunkach suszy. Wizualizacje analizy składowych głównych zostały wykonane za pomocą pakietu stats w R.

Literatura

- [1] Salachna P., Wesołowska A., Zawadzińska A., Kańczurzevska M., Sawikowska A., Daras A., Neugebauerová J., Meller E., Pietrak A., Piechocki R., Łopusiewicz Ł., *Flaxseed oil cake improves basil (*Ocimum basilicum* L.) yield under drought stress by increasing herb biomass and quality of essential oil*, Industrial Crops & Products, 220 (2024), 119295.
- [2] Langfelder P., Horvath S., *WGCNA: an R package for weighted correlation network analysis* BMC Bioinformatics, 9 (2008), 559.
- [3] Piasecka A., Sawikowska A., Kuczyńska A., Ogrodowicz P., Mikołajczak K., Krystowiak K., Gudyś K., Guzy-Wróbelska J., Krajewski P., Kachlicki P., *Drought related secondary metabolites of barley (*Hordeum vulgare* L.) leaves and their association with mQTLs*, Plant J 89 (2017), 898–913.

Analiza układów wierszowo-kolumnowych – model losowy

Idzi Siatkowski

Uniwersytet Przyrodniczy w Poznaniu

W prezentacji przedstawiono analizy statystyczne układów wierszowo-kolumnowych w sytuacji, gdy efekty wierszowe oraz efekty kolumnowe są losowe. Prezentowane metody zilustrowano kilkoma przykładami.

Funkcjonalna analiza wariancji w detekcji stanu technicznego maszyn pracujących przy zmiennym obciążeniu

Maria Skupień

Uniwersytet Komisji Edukacji Narodowej w Krakowie

Diagnozowanie uszkodzeń w maszynach pracujących rotacyjnie, przy zmiennym obciążeniu, jest wymagającym zadaniem. Dotychczasowe metody analizy sygnałów wibracyjnych nadal nie radzą sobie zbyt dobrze z wykrywaniem subtelnych uszkodzeń. Ponieważ amplituda drgań jest zależna od zmiennego obciążenia, w systemach monitorowania maszyn mogą pojawiać się fałszywe alarmy. W związku z tym istnieje zapotrzebowanie na metody diagnostyczne, które są niewrażliwe na zmiany warunków pracy diagnozowanej maszyny. Wychodząc od analizy rzędów, metody wolnej od tych ograniczeń, badany jest wpływ obciążenia na wartości parametrów diagnostycznych za pomocą stanowiska laboratoryjnego. Zebrane sygnały zostały wygładzone, a następnie zastosowano narzędzia analizy danych funkcjonalnych. Testem FANOVA porównano różnice między krzywymi referencyjnymi a krzywymi z potencjalnymi nieprawidłowościami. Różnice te okazały się statystycznie istotne, w przeciwieństwie do standardowych miar. Ze względu na czułość testu FANOVA, zastosowano także inne testy: punktowy test t dla każdej prędkości obrotowej oraz test globalny, który dostarcza zagregowanej informacji, pomocnej w weryfikacji hipotezy zerowej. Metody te umożliwiają wykrycie rzędów z najwyższą różnicą pomiędzy sygnałami z poprawnie pracującej maszyny a tymi z wprowadzonymi uszkodzeniami. Uzyskane wyniki pozwalają na automatyzację detekcji uszkodzeń, w zależności od ich stopnia, w maszynach pracujących przy zmiennym obciążeniu.

Literatura

- [1] Pawlik P., *Analysis of the impact of oil viscosity on the vibroacoustic signal of a planetary gearbox operating at variable load*, Diagnostyka 21, 2020.
- [2] Horváth L., Kokoszka P., *Inference for functional data with applications*, Vol. 200, Springer Science & Business Media, Springer, 2012
- [3] Górecki T., Smaga Ł., *A comparison of tests for the one-way anova problem for functional data*, Computational Statistics, 30 (4), 2015.

Rozkład normalny skośny ze składnikiem uplastyczniającym II

Damian Stoltmann

Uniwersytet Pomorski w Słupsku

Pierwszym (głównym) celem jest wprowadzenie nowego elastycznego rozkładu zdefiniowanego w nieskończonej domenie. Ten rozkład nazywa się rozkładem normalnym skośnym ze składnikiem uplastyczniającym II. Drugim (dodatkowym) celem jest chronologiczny przegląd rozkładów należących do dużej rodziny rozkładów uplastycznionych normalnych. Przedstawiono niektóre właściwości propozycji, takie jak PDF, CDF, kwantyle, generator, momenty, skośność, kurtoza i momenty. Nieznane parametry nowego rozkładu oszacowano metodą największego prawdopodobieństwa. Przedstawiono również entropię Shannona i macierz informacji Fishera. Podano ilustrujące przykłady stosowalności i elastyczności wprowadzonego rozkładu. Podano PDF-y dużej rodziny rozkładów uplastycznionych normalnych.

Skew Bimodal Power Laplace Distribution

Piotr Sulewski

Uniwersytet Pomorski

This article has two goals. The first (main) goal is to introduce a new flexible distribution defined in an infinite domain. This distribution is named the skew bimodal power Laplace distribution. The second (additional) goal is a chronological overview of distributions belonging to the large family of the skew Laplace distributions. Some properties of the proposal such as PDF, CDF, quantiles, generator, moments, skewness, kurtosis and moments of order statistics are presented. The unknown parameters of the new distribution are estimated by the maximum likelihood method. The Shannon entropy and Fisher Information Matrix are also presented. Illustrative examples of applicability and flexibility of the introduced distribution are given. The most important R codes and PDFs of distributions used in real data examples are given in the Appendix.

Literatura

- [1] Hady D.A., *Transmuted Laplace distribution: properties and applications*, American Journal of Applied Mathematics and Statistics, 4(3) (2016), 94-98.
- [2] Nassar M.M., *The Kumaraswamy Laplace Distribution*, Pakistan Journal of Statistics and Operation Research, (2016), 609-624.
- [3] Nekoukhou V., Alamatsaz M.H., *A family of skew-symmetric-Laplace distributions*, Statistical papers, 53, (2012), 685-696.
- [4] Shah S., Hazarika P.J., Chakraborty S., Alizadeh M., *The Balakrishnan-Alpha-Beta-Skew-Laplace Distribution: Properties and Applications*, Statistics, Optimization and Information Computing, 11(3), (2023), 755-772.
- [5] Sulewski P., *Modified Lilliefors goodness-of-fit test for normality*, Communications in Statistics - Simulation and Computation, 51(3) (2020), 1199-1219.
- [6] Tovar-Falon R., Martinez-Florez G. *A New Class of Exponentiated Beta-Skew-Laplace Distribution*, Anais da Academia Brasileira de Ciências, 94 (2022).

O optymalnej adaptacji w poissonowskich problemach odwrotnych

Zbigniew Szkutnik

Akademia Górniczo-Hutnicza w Krakowie

Przedstawiona zostanie nowa, oparta na wygładzonych resztach, postać zasady rozbieżności dla wyboru estymatora w poissonowskich problemach odwrotnych z operatorem zwartym (por. [4]). Dla szerokiej klasy estymatorów opartych na filtrach spektralnych, prowadzi ona do estymatorów o istotnie optymalnych tempach zbieżności w normie L^2 na klasach zadanych przez tzw. warunki źródłowe, jeżeli stopień wygładzania reszt jest odpowiednio powiązany ze (znanymi) własnościami wygładzającymi operatora. Poprawia to istotnie wyniki z [2] dla standardowej zasady rozbieżności, która prowadziła do suboptymalnych temp zbieżności estymatorów. Wyniki teoretyczne będą zilustrowane eksperymentem symulacyjnym dla dwóch problemów stereologicznych: problemu Wicksella (z operatorem Abela) i problemu Spektora-Lorda-Willisa (z operatorem całkowania). Problem poissonowski będzie przedstawiony w szerszym kontekście problemów odwrotnych z białym szumem.

Ogólna idea wygładzania reszt została zaczerpnięta z [3], a użyte techniki analityczne są podobne do tych użytych w [1] i [2]. Dodatkowo, pożyteczne okazało się wprowadzenie i użycie nowych pojęć, np. tzw. γ -ograniczoności filtra spektralnego.

Literatura

- [1] Harrach B., Jahn T., Potthast R., *Beyond the Bakushinskii veto: Regularising linear inverse problems without knowing the noise distribution*, Numerische Mathematik, 145 (2020), 581–603.
- [2] Mika G., Szkutnik Z., *Towards adaptivity via a new discrepancy principle for Poisson inverse problems*, Electronic Journal of Statistics, 15 (2021), 2029-2059.
- [3] Stankewitz B., *Smoothed residual stopping for statistical inverse problems via truncated SVD estimation*, Electronic Journal of Statistics, 14 (2020), 3396-3428.
- [4] Szkutnik Z., *Weighted discrepancy principle and optimal adaptivity in Poisson inverse problems*, ALEA, 21 (2024), 1101-1122.

Kluczowe rozkłady życia komponentów gwarantujące systemowi dziedziczenie wybranych własności starzenia

Jorge Navarro¹, Tomasz Rychlik², Magdalena Szymkowiak³

¹ Universidad de Murcia

² Instytut Matematyczny PAN

² Politechnika Poznańska

W pracy pokazano, że niektóre rozkłady odgrywają kluczową rolę w zachowaniu własności starzenia przy budowie układów koherentnych (lub mieszanych). Jeśli dla kluczowego rozkładu jakaś własność czasów życia komponentów jest dziedziczona przez system, to jest ona także zachowana dla wszystkich rozkładów o tej własności.

W głównych klasach starzenia, tj. IFR/DFR (rosnąca/malejąca intensywność uszkodzeń), IFRA/DFRA (rosnąca/malejąca średnia intensywność uszkodzeń), NBU/NWU (nowy lepszy/gorszy niż używany), rozkład wykładniczy jest kluczowy, ponieważ charakteryzuje komponenty nie starzejące się (z brakiem pamięci). W innych klasach, np. ID/DD (rosnąca/malejąca gęstość), kluczową rolę odgrywa rozkład jednostajny. W pracy wykorzystano reprezentację funkcji niezawodności systemu przy pomocy jego funkcji zniekształcającej. Otrzymane wyniki dotyczące zachowania klas starzenia można zastosować do systemów o różnych strukturach, z niezależnymi lub zależnymi komponentami o tych samych rozkładach życia.

Literatura

- [1] Arnold B.C., Rychlik T., Szymkowiak M. *Preservation of distributional properties of component lifetimes by system lifetimes*, *Test*, 31 (2022) 901–930.
- [2] Navarro J. *Introduction to System Reliability Theory*. Springer, 2022.
- [3] Rychlik T., Szymkowiak M. *Signature conditions for distributional properties of system lifetimes if component lifetimes are iid exponential*, *IEEE Transactions on Reliability*, 71 (2022), 590–602.

Metody klasyfikacji w sytuacji przesunięcia rozkładu warunkowego zmiennej odpowiedzi między zbiorem treningowym i testowym

Paweł Teisseyre^{1,2}, Jan Mielniczuk^{1,2}

¹ Wydział Matematyki i Nauk Informatycznych, Politechnika Warszawska

² Instytut Podstaw Informatyki, PAN

W wielu praktycznych zastosowaniach uczenia maszynowego pojawia się rozbieżność między rozkładem prawdopodobieństwa, z którego generowane są dane treningowe, a rozkładem z którego pochodzą dane testowe. Nieuwzględnienie przesunięcia rozkładów przy dopasowaniu modelu klasyfikacji może prowadzić do istotnego obniżenia jakości predykcji. W literaturze rozważa się dwa scenariusze: CS (ang. covariate shift) [1] i LS (ang. label shift) [2]. W schemacie CS zmienia się rozkład brzegowy cech, zaś rozkład warunkowy zmiennej odpowiedzi pod warunkiem cech pozostaje niezmienny. W schemacie LS, zmienia się rozkład apriori zmiennej odpowiedzi, zaś rozkład cech, przy ustalonej wartości zmiennej odpowiedzi pozostaje niezmienny. Powyższe schematy nie obejmują wszystkich możliwych przypadków przesunięcia rozkładów. W pracy [3] zaproponowano nowy schemat SJS (ang. Sparse Joint Shift), będący uogólnieniem CS i LS, w którym zakłada się, że zmienia się rozkład zmiennej odpowiedzi oraz wybranych cech. W referacie przedstawimy nowy schemat CPS (ang. Conditional Probability Shift) [4], obejmujący najbardziej interesujący przypadek SJS, w którym zmienia się rozkład warunkowy zmiennej odpowiedzi przy ustalonych wartościach wybranych cech. Przedstawimy model dla przesunięcia warunkowego prawdopodobieństwa, bazujący na regresji wielomianowej. Zaproponowaliśmy algorytm EM (ang. Expectation Maximization), który umożliwia estymację parametrów w takim modelu. Eksperymenty przeprowadzone na syntetycznych zbiorach danych oraz na bazie medycznej MIMIC pokazują, że proponowana metoda pozwala uzyskać większą dokładność klasyfikacji w porównaniu do metod konkurencyjnych bazujących na założeniach LS i SJS. Co istotne, metoda może być stosowana w połączeniu z dowolnym modelem klasyfikacji.

Literatura

- [1] Bickel, S., Bruckner, M., Scheffer, T., *Discriminative learning under covariate shift*, Journal of Machine Learning Research 10 (75), 2137–2155, 2009.
- [2] Lipton, Z.C., Wang, Y., Smola, A.J., *Detecting and correcting for label shift with black box predictors*, Proceedings of the 35th International Conference on Machine Learning ICML' 18, 3128–3136, 2018.
- [3] Chen, L., Zaharia, M., Zou, J., *Estimating and explaining model performance when both covariates and labels shift*, Proceedings of the 36th International Conference on Neural Information Processing Systems NIPS' 22, 1–13, 2024.
- [4] Teisseyre, P., Mielniczuk, J., *Probabilistic classification when conditional distributions of labels for source and target domains differ*, w recenzji, 2024.

Model klasyfikacyjny dla spraw sądowych Banku

Łukasz Waszak

Alior Bank SA

W działalności bankowej, na potencjalnie przegrane sprawy sporne, banki muszą tworzyć rezerwy w swoich sprawozdaniach finansowych. W momencie prawomocnego rozstrzygnięcia sprawy sądowej taka rezerwa staje się stratą rzeczywistą banku. W związku z tym, rezerwy banku powinny być na adekwatnym poziomie. Alior Bank zbudował narzędzie do klasyfikacji takich spraw (wygrana / przegrana). W trakcie referatu zaprezentuję ogólną ideę podejścia do budowy takiego modelu w praktyce. Następnie, pokrótce przedstawię techniki uczenia maszynowego ML użyte w modelu. Omówię również różne miary oceny i skuteczności tak zbudowanego modelu.

Literatura

- [1] L. Breiman, *Random Forests*, Machine Learning, 45, 5-32, 2001.
- [2] L. Breiman, A. Cutler, *Extremely randomized trees*, Machine Learning, 63, 3-42, 2006.
- [3] L. Breiman, J. Friedman, R. Olshen, C. Stone, *Classification and Regression Trees*, Wadsworth, Belmont, CA, 1984.
- [4] T. Hastie, R. Tibshirani, J. Friedman, *Elements of Statistical Learning*, Springer, 2009.
- [5] A. Ogłoszka, Ł. Smaga, *Classification methods in the diagnosis of breast cancer*, Biometrical Letters 59, 2022.

Semi-supervised learning with semi-continuous hidden Markov models

Filip Wichrowski, Katarzyna Kaczmarek-Majer

Instytut Badań Systemowych PAN

Hidden Markov Models (HMMs) are widely used for modeling sequential data in fields such as speech recognition, bioinformatics, and time series analysis. A key challenge in HMMs is parameter estimation, which is typically performed using the Baum-Welch algorithm, an Expectation-Maximization (EM) approach. However, the algorithm's performance is highly sensitive to initial parameter values, leading to local optima and sub-optimal model performance. In this article, we focus on addressing this issue within the framework of semi-supervised learning, where partially labeled data is available. On three benchmarks we have constructed, we compare different strategies for initialization of the HMM parameters in the Baum-Welch algorithm, such as random initialization and initialization based on the k-Means algorithm. We show that understanding the structure of the transitions between hidden states can drastically improve model's performance, especially when there are a lot of missing labels in the data. Finally, we compare the resulting models with popular semi-supervised classifiers, such as Label Propagation, Label Spreading and semi-supervised SVM. The results suggest that (a) the proposed initialization strategies can significantly enhance the performance of HMMs in semi-supervised settings, providing a practical solution to a long-standing challenge in parameter estimation and (b) HMMs offer an excellent alternative that is often superior to other popular classifiers.

Literatura

- [1] Huggins-Daines D., Rudnicky A. I., *A constrained baum-welch algorithm for improved phoneme segmentation and efficient training*, Proc. Interspeech, 2016, ss. 1205-1208.
- [2] X.D. Huang, M.A. Jack, *Semi-continuous hidden Markov models for speech signals*, Computer Speech & Language, 1989, ss. 239-251.
- [3] Rabiner L., Juang B., *An introduction to hidden Markov models*, IEEE ASSP Magazine, 1986, ss. 4-16.

Test permutacyjny w problemie k -prób dla danych prawostronnie cenzurowanych

Grzegorz Wyłupek

Uniwersytet Wrocławski

W referacie przedstawimy nowe rozwiązanie nieparametrycznego problemu k -prób dla danych prawostronnie cenzurowanych. Zastosowane podejście polega na połączeniu efektywnej statystyki wynikowej z selekcją modelu. W pierwszym kroku dokonujemy reparametryzacji problemu do k dwupróbkowych podproblemów, w których porównujemy l -tą próbę z połączonymi pozostałymi próbami. Następnie, w każdym podproblemie statystyką pomocniczą jest suma kwadratów standaryzowanych ważonych statystyk log-rank, powiązanych z układem wielomianów Laguerre. Wprowadzona reguła selekcji modelu, decyduje o liczbie istotnych ważonych statystyk log-rank włączanych do pomocniczej dwupróbkowej statystyki. Ich optymalna kombinacja liniowa prowadzi do finałowej statystyki testowej. Przy prawdziwości hipotezy zerowej, statystyka testowa ma asymptotyczny rozkład chi-kwadrat z $k - 1$ stopniami swobody, a odpowiadający jej test jest zgodny dla szerokiej klasy alternatyw. Ponieważ dane są prawostronnie cenzurowane, to, aby kontrolować błąd pierwszego rodzaju, konieczna jest konstrukcja testu warunkowego. W tym celu wykorzystujemy podejście permutacyjne. Przeprowadzone obszerne badania symulacyjne pokazują bardzo dobre własności nowego testu.

Literatura

- [1] Bagdonavičius, V. B., Levulienė, R. J., Nikulin, M. S., Zdorova-Cheminade, O., *Tests for equality of survival distributions against non-location alternatives*, Lifetime Data Analysis, 10 (2004), 445-460.
- [2] Breslow, N., *A generalized Kruskal-Wallis test for comparing K samples subject to unequal patterns of censorship*, Biometrika, 57 (1970), 579-594.
- [3] Chen, Z., Huang, H., Qiu, P., *Comparison of multiple hazard rate functions*, Biometrics, 72 (2016), 39-45.
- [4] Gorfine, M., Schlesinger, M., Hsu, L., *K -sample omnibus non-proportional hazards tests based on right-censored data*, Statistical Methods in Medical Research, 29 (2020), 2830-2850.
- [5] Liu, Y., Yin, G., *Partitioned log-rank tests for the overall homogeneity of hazard rate functions*, Lifetime Data Analysis, 23 (2017), 400-425.
- [6] Martínez-Camblor, P., *Comparing k -independent and right censored samples based on the likelihood ratio*, Computational Statistics, 25 (2010), 363-374.
- [7] Wyłupek, G., *A permutation test for the two-sample right-censored model*, Annals of the Institute of Statistical Mathematics, 73 (2021), 1037-1061.
- [8] Wyłupek, G., *A permutation k -sample test for right-censored data*, w przygotowaniu (2024).

Przedziały ufności dla „wrażliwej proporcji” w badaniach ankietowych

Marta Zalewska¹, Wojciech Niemiro^{2,3}

¹ Warszawski Uniwersytet Medyczny

² Uniwersytet Warszawski

³ Uniwersytet Mikołaja Kopernika, Toruń

Badanie cech wrażliwych jest ważnym problemem w statystyce ankietowej (*survey statistics*). Chodzi tu przede wszystkim o zachowania nieakceptowane społecznie, na przykład używanie narkotyków, korupcję, przerywanie ciąży itp. Aby uzyskać prawdziwe odpowiedzi od respondentów, opracowano kilka metod konstrukcji ankiet, w tym technikę liczenia/sumowania elementów (ICT/IST – *Item Count / Sum Technique*) [4]. W przypadku binarnej cechy wrażliwej celem badania jest proporcja osób z cechą wrażliwą w populacji. Estymatory tej proporcji są zwykle uzupełnione standardowymi przedziałami ufności opartymi na przybliżeniu normalnym. Jednak w praktyce rzeczywiste prawdopodobieństwo pokrycia dla takich przedziałów nie jest wystarczająco bliskie nominalnemu poziomowi ufności.

Proponujemy dwie nowe metody konstruowania przedziałów ufności dla proporcji cechy wrażliwej w badaniach ankietowych typu ICT/IST. Jedna z tych metod wykorzystuje parametryczny bootstrap i algorytm EM (*Expectation-Maximalization*) [2]. Druga metoda opiera się na zupełnie innym pomysśle, mianowicie na obliczeniu dokładnego przedziału ufności w modelu parametrycznym, a następnie podstawieniu estymatora w miejsce parametru zakłócającego [1]. Używamy rozkładu Poissona do modelowania odpowiedzi na pytanie neutralne maskujące, zgodnie z [3].

Literatura

- [1] Stanisław Jaworski, Wojciech Zieliński. *The optimal sample size in the crosswise model for sensitive questions*. *Applicationes Mathematicae* (2023) 50, 21-34 <https://doi.org/10.4064/am2486-11-2023>
- [2] Barbara Kowalczyk, Wojciech Niemiro, Robert Wieczorkowski. *Item count technique with a continuous or count control variable for analyzing sensitive questions in surveys*. *Journal of Survey Statistics and Methodology* (2023) 11, 4, 919–941, <https://doi.org/10.1093/jssam/smab043>
- [3] G-L. Tian, M-L. Tang, Q. Wu, Y. Liu. *Poisson and negative binomial item count techniques for surveys with sensitive question*. *Statistical Methods in Medical Research* (2017) 26, 931-947.
- [4] M. Trappman, I. Krumpal, A. Kirchner and B. Jann. *Item Sum: A New Technique for Asking Quantitative Sensitive Questions*. *Journal of Survey Statistics and Methodology* (2014) 2, 58-77.

Lista uczestników

1. **Mgr Konrad Banaś**
Uniwersytet Przyrodniczy w Poznaniu
Katedra Metod Matematycznych i Statystycznych
konrad.banas@up.poznan.pl
2. **Dr hab. Mariusz Bieniek**
Uniwersytet Marii Curie-Skłodowskiej w Lublinie
Instytut Matematyki
mariusz.bieniek@umcs.lublin.pl
3. **Prof. Małgorzata Bogdan**
Uniwersytet Wrocławski
Instytut Matematyczny
malgorzata.bogdan20@gmail.com
4. **Dr Jacek Bojarski**
Uniwersytet Zielonogórski
Instytut Matematyki
j.bojarski@im.uz.zgora.pl
5. **Dr Anna Cena**
Politechnika Warszawska
Wydział Matematyki i Nauk Informatycznych
anna.cena@pw.edu.pl
6. **Mgr Adam Chojecki**
Politechnika Warszawska
Wydział Matematyki i Nauk Informatycznych
adam.prze.choj@gmail.com
7. **Dr Bogdan Ćmiel**
Akademia Górniczo-Hutnicza w Krakowie
Wydział Matematyki Stosowanej
cmielbog@gmail.com
8. **Iza Danielewska**
Politechnika Warszawska
Wydział Matematyki i Nauk Informatycznych
01161562@pw.edu.pl
9. **Dr hab. Łukasz Dębowski**
Instytut Podstaw Informatyki PAN
ldebowsk@ipipan.waw.pl

10. **Prof. Paweł Dłotko**
Instytut Matematyczny PAN
Dioscuri Centre in Topological Data Analysis
pdlotko@impan.pl
11. **Dr hab. Anna Dudek**
Akademia Górniczo-Hutnicza w Krakowie
Wydział Matematyki Stosowanej
aedudek@agh.edu.pl
12. **Dr hab. Katarzyna Filipiak**
Politechnika Poznańska
Instytut Matematyki
katarzyna.filipiak@put.poznan.pl
13. **Dr hab. Konrad Furmańczyk**
Szkoła Główna Gospodarstwa Wiejskiego w Warszawie
Instytut Informatyki Technicznej
konrad_furmanczyk@sggw.edu.pl
14. **Dr hab. Marek Gągolewski**
Politechnika Warszawska
Wydział Matematyki i Nauk Informacyjnych
oraz Instytut Badań Systemowych PAN
marek.gagolewski@pw.edu.pl
15. **Dr hab. Agnieszka Goroncy**
Uniwersytet Mikołaja Kopernika w Toruniu
Wydział Matematyki i Informatyki
gemini@mat.umk.pl
16. **Prof. Przemysław Grzegorzewski**
Politechnika Warszawska
Wydział Matematyki i Nauk Informacyjnych
oraz Instytut Badań Systemowych PAN
przemyslaw.grzegorzewski@pw.edu.pl
17. **Mgr Aleksandra Grzelik**
Instytut Badań Systemowych PAN
agrzelik@ibspan.waw.pl
18. **Prof. Szymon Jaroszewicz**
Politechnika Warszawska
Wydział Matematyki i Nauk Informacyjnych
oraz Instytut Podstaw Informatyki PAN
s.jaroszewicz@ipipan.waw.pl
19. **Dr Krzysztof Jasiński**
Uniwersytet Mikołaja Kopernika w Toruniu
Wydział Matematyki i Informatyki
krzys@mat.umk.pl

20. **Dr hab. Alicja Jokiel-Rokita**
Politechnika Wrocławska
Wydział Matematyki
alicja.jokiel-rokita@pwr.edu.pl
21. **Mgr Marta Kańczurzevska-Staszak**
Politechnika Poznańska
Instytut Matematyki
marta.kanczurzevska-staszak@put.poznan.pl
22. **Dr hab. Bartosz Kołodziejek**
Politechnika Warszawska
Wydział Matematyki i Nauk Informatycznych
bartosz.kolodziejek@pw.edu.pl
23. **Mgr Anna Kozak**
Politechnika Warszawska
Wydział Matematyki i Nauk Informatycznych
anna.kozak@pw.edu.pl
24. **Dr hab. Łukasz Lenart**
Uniwersytet Ekonomiczny w Krakowie
Katedra Matematyki
lenartl@uek.krakow.pl
25. **Prof. Ryszard Magiera**
Politechnika Wrocławska
Wydział Matematyki
ryszard.magiera@pwr.edu.pl
26. **Mgr Bartosz Majewski**
Akademia Górniczo-Hutnicza w Krakowie
Wydział Matematyki Stosowanej
bartmaje@agh.edu.pl
27. **Mgr Karolina Marek**
Akademia Górniczo-Hutnicza w Krakowie
Wydział Matematyki Stosowanej
karolinm@agh.edu.pl
28. **Dr hab. Błażej Miasojedow**
Uniwersytet Warszawski
Wydział Matematyki, Informatyki i Mechaniki
bmiasojedow@gmail.com
29. **Prof. Jan Mielniczuk**
Politechnika Warszawska
Wydział Matematyki i Nauk Informatycznych
oraz Instytut Podstaw Informatyki PAN
miel@pipan.waw.pl
30. **Mgr Malwina Mrowińska**
Politechnika Poznańska
Instytut Matematyki
malwina.mrowinska@put.poznan.pl

31. **Prof. Wojciech Niemi**
Uniwersytet Warszawski
Wydział Matematyki, Informatyki i Mechaniki
wniem@mimuw.edu.pl
32. **Dr hab. Karol Opara**
Polska Akademia Nauk PAN
karol.opara@ibspan.waw.pl
33. **Mgr Jagoda Papis**
Politechnika Poznańska
Instytut Matematyki
jagoda.papis@put.poznan.pl
34. **Mgr Sylwester Piątek**
Politechnika Wrocławska
Wydział Matematyki
sylwester.piatek@pwr.edu.pl
35. **Dr Agnieszka Piliszek**
Politechnika Warszawska
Wydział Matematyki i Nauk Informatycznych
apiliszek@gmail.com
36. **Mgr Katarzyna Pypka**
Uniwersytet Zielonogórski
katarzynapypka@gmail.com
37. **Dr hab. Wojciech Rejchel**
Uniwersytet Mikołaja Kopernika w Toruniu
Wydział Matematyki i Informatyki
wrejchel@gmail.com
38. **Dr Krzysztof Rudaś**
Politechnika Warszawska
Wydział Matematyki i Nauk Informatycznych
oraz Instytut Podstaw Informatyki PAN
krzysztof.rudas@pw.edu.pl
39. **Mgr Jakub Rutkowski**
Akademia Górniczo-Hutnicza w Krakowie
Wydział Matematyki Stosowanej
jrutkowski@agh.edu.pl
40. **Prof. Tomasz Rychlik**
Instytut Matematyczny PAN
trychlik@impan.pl
41. **Dr hab. Aneta Sawikowska**
Uniwersytet Przyrodniczy w Poznaniu
Katedra Metod Matematycznych i Statystycznych
aneta.sawikowska@up.poznan.pl

42. **Prof. Idzi Siatkowski**
Uniwersytet Przyrodniczy w Poznaniu
Katedra Metod Matematycznych i Statystycznych
idzi.siatkowski@up.poznan.pl
43. **Mgr Maria Skupień**
Uniwersytet Komisji Edukacji Narodowej w Krakowie
Instytut Matematyki
maria.skupien@uken.krakow.pl
44. **Mgr Damian Stoltmann**
Uniwersytet Pomorski w Słupsku
Instytut Nauk Ścisłych i Technicznych
damian.stoltmann@upsl.edu.pl
45. **Dr hab. Piotr Sulewski**
Uniwersytet Pomorski w Słupsku
Instytut Nauk Ścisłych i Technicznych
piotr.sulewski@upsl.edu.pl
46. **Prof. Zbigniew Szkutnik**
Akademia Górniczo-Hutnicza w Krakowie
Wydział Matematyki Stosowanej
szkutnik@agh.edu.pl
47. **Dr hab. Magdalena Szymkowiak**
Politechnika Poznańska
Instytut Automatyki i Robotyki
magdalena.szymkowiak@put.poznan.pl
48. **Dr hab. Paweł Teisseyre**
Wydział Matematyki i Nauk Informatycznych
oraz Instytut Podstaw Informatyki PAN
teisseyrep@ipipan.waw.pl
49. **Dr Łukasz Waszak**
Alior Bank SA
lukasz.waszak@alior.pl
50. **Mgr Filip Wichrowski**
Instytut Badań Systemowych PAN
fwichrow@ibspan.waw.pl
51. **Dr Grzegorz Wyłupek**
Uniwersytet Wrocławski
Instytut Matematyczny
wylupek@math.uni.wroc.pl
52. **Dr hab. Marta Zalewska**
Warszawski Uniwersytet Medyczny
Zakład Profilaktyki Zagrożeń Środowiskowych, Alergologii i Immunologii
marta.zalewska@wum.edu.pl

Indeks

- Banaś Konrad, 19
Bertail Patrice, 27, 41
Bieniek Mariusz, 20, 52
Bogdan Małgorzata, 21
Bojarski Jacek, 22
Bujak Henryk, 19
- Chojecki Adam, 21, 23
Chrzanowska Mariola, 29
- Ćmiel Bogdan, 24
- Danielewska Iza, 25
Dębowski Łukasz, 26
Dłotko Paweł, 16
Drapieński Dawid, 22
Dudek Anna, 27, 40, 41, 51
- Filipiak Katarzyna, 28, 44
Furmańczyk Konrad, 29, 43
- Gągolewski Marek, 14
Goroncy Agnieszka, 30, 34
Graczyk Piort, 23
Grzegorzewski Przemysław, 31
Grzelik Aleksandra, 32
- Hejny Ivan, 21
- Ishi Hideyuki, 23
Iskrzyński Mateusz, 32
- Jaroszewicz Szymon, 33, 50
Jasiński Krzysztof, 30, 34
Jokiel-Rokita Alicja, 35, 47
- Kaczmarek-Majer Katarzyna, 62
Kańczurzevska-Staszak Marta, 36, 53
Klein Daniel, 44
Kołodziejek Bartosz, 21, 23, 25, 37
- Korejwo Faustyna, 30
Kozak Anna, 38
- Ledwina Teresa, 24
Lenart Łukasz, 39
Lenartowicz Tomasz, 19
- Łącka Agnieszka, 19
- Magiera Ryszard, 35
Majewski Bartosz, 40
Marek Karolina, 41
Mazur Stepan, 44
Miasojedow Błażej, 42
Mielniczuk Jan, 43, 49, 60
Mrowińska Malwina, 44
- Napolitano Antonio, 51
Navarro Jorge, 59
Niedziela Maciej, 22
Niemiro Wojciech, 29, 64
- Opara Karol, 45
Osiecka Agnieszka, 19
- Papis Jagoda, 46
Piątek Sylwester, 47
Pieszko Konrad, 48
Przystalski Marcin, 19
Puchalska Aleksandra, 32
Pypka Katarzyna, 22, 48
- Rejchel Wojciech, 49
Ruczyński Hubert, 38
Rudaś Krzysztof, 33, 50
Rudzate Marta, 30
Rutkowski Jakub, 51
Rychlik Tomasz, 46, 52, 59
- Sawikowska Aneta, 36, 53

Siatkowski Idzi, 54
Skupień Maria, 55
Stein Jarosław, 22
Stoltmann Damian, 56
Sulewski Piotr, 57
Szkutnik Zbigniew, 58
Szymkowiak Magdalena, 46, 59

Teisseyre Paweł, 43, 49, 60
Topolnicki Rafał, 47

Wallin Jonas, 21
Waszak Łukasz, 61

Wesołowski Jacek, 25
Wichrowski Filip, 62
Wojciech Magdalena, 48
Wojdyła Jakub, 51
Wyłupek Grzegorz, 63

Zaborniak Gabriel, 22
Zalewska Marta, 29, 64
Zeng Xiaolin, 25
Ziembicki Piotr, 22

Żuber Kornel, 22