XXXIX Konferencja

STATYSTYKA MATEMATYCZNA WISŁA 2013

Wisła, 2-6 grudnia 2013

PAN
POLSKA AKADEMIA NAUK

Politechnika Wrocławska
Instytut Matematyki i Informatyki

Organizatorzy konferencji

Komisja Statystyki Matematycznej Komitetu Matematyki PAN
Instytut Matematyki i Informatyki PWr

Komitet organizacyjny

Małgorzata Bogdan – przewodnicząca
Alicja Janic
Monika Kaczmarz
Piotr Sobczyk
Krzysztof Szajowski
Piotr Szulc

Miejsce konferencji

Hotel Gawra
Wisła-Uzdrowisko
ul. Górnośląska 44
43-360 Wisła

Redakcja

Jerzy Baran
Piotr Szulc

Druk i oprawa

Drukarnia Oficyny Wydawniczej PWr

Część I

Program konferencji

Poniedziałek, 2 grudnia 2013

08:00	**Śniadanie**
09:00	Otwarcie konferencji
09:15 - 10:00	**Tomasz Burzykowski** Wprowadzenie do genetyki i zastosowań statystyki w genetyce
10:00 - 10:30	**Przerwa**
10:30 - 10:50	**Stanisław Mejza** Profesor Tadeusz Caliński
10:50 - 11:10	Krystyna Katulska, **Łukasz Smaga** KONKURS D-efektywność chemicznych układów wagowych przy skorelowanych błędach losowych
11:10 - 11:30	**Piotr Szulc** KONKURS Skąd się biorą „gorące" miejsca na genomie?
11:30 - 11:50	**Przerwa**
11:50 - 12:10	**Aleksandra Maj-Kańska**, Piotr Pokarowski, Agnieszka Prochenka KONKURS Wybór modelu liniowego poprzez jednoczesne usuwanie zmiennych ciągłych i łączenie poziomów zmiennych czynnikowych
12:10 - 12:30	**Hubert Szymanowski** KONKURS O normowanym estymatorze Dantziga
12:30 - 12:50	Pin Pin Oh, **Karol Opara** KONKURS Wklęsłe funkcje straty w kinetycznym modelowaniu reakcji chemicznych
13:00	**Obiad**
15:00 - 15:20	Mirosław Krzyśko, **Łukasz Waszak** KONKURS Analiza korelacji kanonicznych dla wielozmiennych danych funkcjonalnych
15:20 - 15:40	Tomasz Rychlik, **Patryk Miziuła** KONKURS Oszacowania wariancji czasu życia systemów niezawodnościowych z permutowalnymi elementami
15:40 - 16:00	**Piotr Sobczyk** KONKURS Rozpoznawanie sekwencji kodujących białko w genomach prokariotycznych za pomocą ukrytych łańcuchów Markowa
16:00 - 16:30	**Przerwa**
16:30 - 16:50	**Krzysztof Szajowski** Wykrywanie losowej liczby rozregulowań z zadaną precyzją
16:50 - 17:10	**Agnieszka Stępień-Baran** Estymacja sekwencyjna parametru położenia i potęg parametru skali
17:10 - 17:30	**Anna Czapkiewicz**, Paweł Jamer Estymacja parametrów przełącznikowych modeli wielowymiarowych szeregów czasowych
18:00	**Kolacja**
19:00	**Otwarta dyskusja**

Wtorek, 3 grudnia 2013

08:00	**Śniadanie**
09:00 - 09:45	**Tomasz Burzykowski** Przegląd problemów i metod związanych z analizą danych dotyczących ekspresji genów
09:45 - 10:30	**Przerwa**
10:30 - 10:50	**Teresa Ledwina**, Grzegorz Wyłupek Wnioskowanie o dodatniej kwadrantowej zależności, I
10:50 - 11:10	Teresa Ledwina, **Grzegorz Wyłupek** Wnioskowanie o dodatniej kwadrantowej zależności, II
11:10 - 11:30	**Kamil Dyba** Testowanie wielowymiarowej stochastycznej dominacji z wykorzystaniem wielowymiarowych funkcji kwantylowych
11:30 - 11:50	**Przerwa**
11:50 - 12:10	**Przemysław Grzegorzewski**, Hubert Szymanowski Testy zgodności dla nieprecyzyjnych danych
12:10 - 12:30	Tadeusz Inglot, **Dawid Kujawa** Adaptacyjne testy symetrii wokół znanego środka
12:30 - 12:50	Tadeusz Inglot, **Alicja Janic** Adaptacyjne testy wynikowe symetrii wokół nieznanego środka
13:00	**Obiad**
15:00 - 15:20	Rie Enomoto, **Zofia Hanusz**, Kazuyuki Koizumi, Takashi Seo Symulacyjne badanie testów wielowymiarowej normalności opartych na skośności i kurtozie
15:20 - 15:40	**Agnieszka Kulawik** Odporna estymacja parametrów w wielowymiarowym modelu normalnym – wyniki symulacji komputerowych
15:40 - 16:00	Marta Molińska-Glura, Krzysztof Moliński, **Maciej Szydłowski** Analiza skupień w detekcji genów warunkujących cechy ilościowe
16:00 - 16:30	**Przerwa**
16:30 - 16:50	Katarzyna Brzozowska-Rup, **Antoni Leon Dawidowicz** Zastosowanie algorytmu filtru cząsteczkowego do zagadnień ekologii
16:50 - 17:10	**Katarzyna Steliga** Zmodyfikowane rozkłady prawdopodobieństwa typu j i ich charakterystyki
17:10 - 17:30	**Wojciech Zieliński** Najkrótsze przedziały ufności dla prawdopodobieństwa sukcesu w modelu ujemnym dwumianowym
18:00	**Kolacja**

Środa, 4 grudnia 2013

08:00	**Śniadanie**
08:45 - 14:15	**Wycieczka**
14:30	**Obiad**
15:30 - 15:50	**Wojciech Rejchel** Estymatory regresji rangowej oparte na metodzie LASSO
15:50 - 16:10	**Małgorzata Bogdan** SLOPE – nowa procedura identyfikacji istotnych zmiennych w modelach liniowych
16:10 - 16:30	**Paweł Teisseyre** Wykorzystanie metod losowych podprzestrzeni do predykcji i selekcji zmiennych
16:30 - 17:00	**Przerwa**
17:00 - 17:20	**Małgorzata Graczyk** O pewnych własnościach układów wagowych
17:20 - 17:40	Katarzyna Filipiak, **Augustyn Markiewicz** Optymalne układy doświadczalne w modelach współoddziaływania
17:40 - 18:00	Katarzyna Ambroży, **Iwona Mejza** O pewnej metodzie konstrukcji niekompletnych ortogonalnie rozszerzonych układów doświadczalnych typu split-split-plot
18:00	**Kolacja**
19:00	**Posiedzenie Komisji Statystyki Matematycznej Komitetu Matematyki PAN**

Czwartek, 5 grudnia 2013

08:00	**Śniadanie**
09:00 - 09:45	**Tomasz Burzykowski** Przegląd problemów i metod związanych z analizą danych dotyczących ekspresji białek
09:45 - 10:30	**Przerwa**
10:30 - 10:50	**Jan Mielniczuk** Wokół twierdzenia Ruuda
10:50 - 11:10	**Zbigniew Szkutnik** Minimalizacja ryzyka empirycznego w problemach odwrotnych
11:10 - 11:30	**Mariusz Bieniek** Progresywnie modyfikowane statystyki porządkowe
11:30 - 11:50	**Przerwa**
11:50 - 12:10	**Andrzej Michalski** O testowaniu hipotez liniowych w wielowymiarowych modelach liniowych
12:10 - 12:30	**Mariusz Grządziel** Estymacja metodą największej wiarogodności w modelu liniowym z dwoma komponentami wariancyjnymi
12:30 - 12:50	**Czesław Stępniak** Porównywanie układów blokowych ze względu na estymację kwadratową
13:00	**Obiad**
15:00 - 15:20	**Jolanta Grala-Michalak** Redukcja wymiarów danych za pomocą transformacji arctg
15:20 - 15:40	**Katarzyna Stąpor** Diagonalna analiza dyskryminacyjna
15:40 - 16:00	Maria Iwińska, **Magdalena Szymkowiak** Charakteryzacja rozkładów za pomocą wybranych funkcji z teorii niezawodności
16:00 - 16:30	**Przerwa**
16:30 - 16:50	**Krzysztof Jasiński** Maksymalna wariancja k-tych rekordów
16:50 - 17:10	**Maria Kamińska-Zabierowska** Niezmienniczość porządku transformaty GTTT
17:10 - 17:30	**Marcin Makowski**, Edward W. Piotrowski Estymacja parametru rozkładu statystycznego metodą maksymalizacji intensywności przetwarzania informacji
18:30	**Uroczysta kolacja**

Piątek, 6 grudnia 2013

08:00	**Śniadanie**
09:00 - 09:45	**Tomasz Burzykowski** Szukanie QTL (quantitative-trait loci) z zastosowaniem technik sekwencjonowania genów nowej generacji
09:45 - 10:30	**Przerwa**
10:30 - 10:50	**Iwona Żerda** Geometryczna zbieżność algorytmów Gibbsa
10:50 - 11:10	**Anna Szczepańska-Álvarez** Wybór punktów czasowych i układu blokowego w modelu krzywych wzrostu
11:10 - 11:30	**Agnieszka Prochenka** Modelowanie wieku pacjentów za pomocą częstości metylacji cytozyny w DNA
11:30	Zakończenie konferencji
13:00	**Obiad**

Część II

Streszczenia

O pewnej metodzie konstrukcji niekompletnych ortogonalnie rozszerzonych układów doświadczalnych typu split-split-plot

Katarzyna Ambroży, Iwona Mejza

Katedra Metod Matematycznych i Statystycznych
Uniwersytet Przyrodniczy w Poznaniu

W pracy rozważono sytuację, w której układy doświadczalne typu split-split-plot są niekompletne ze względu na obiekty czynników występujących na zagnieżdżonych jednostkach I i II rzędu. Dodatkowo założono, że każdy z nich ma obiekt zwany standardem. Trzeci czynnik, którego obiekty są losowo rozmieszczane na jednostkach III rzędu, występuje w podukładzie ortogonalnym. Metodę konstrukcji oparto na iloczynie Kroneckera macierzy. Natomiast jako układy generujące wybrano te same lub różne układy blokowe z klasy ortogonalnie rozszerzonych częściowo zrównoważonych pod względem efektywności układów blokowych z co najwyżej $(m+1)$-klasami efektywności. Przedstawiono algebraiczne i statystyczne właściwości uzyskanych układów i przykład numeryczny.

Literatura

[1] K. Ambroży, I. Mejza, *Doświadczenia trójczynnikowe z krzyżową i zagnieżdżoną strukturą poziomów czynników,* Wyd. Polskie Towarzystwo Biometryczne and PRODRUK, Poznań, 2006.

[2] K. Ambroży, I. Mejza, *On the efficiency of some non-orthogonal split-plot split-block designs with control treatments,* Journal of Statistical Planning and Inference, Vol. 142, Issue 3, pp. 752-762, 2012.

[3] K. Ambroży, I. Mejza, *A method of constructing incomplete split-split-plot designs supplemented by whole plot and subplot standards and their analysis,* Colloquium Biometricum 43, pp. 59-72, 2013.

[4] T. Caliński, S. Kageyama, *Block Designs. A Randomization Approach, Volume I. Analysis,* Lecture Notes in Statistics 150, Springer-Verlag, New York, 2000.

[5] A.K. Nigam, P.D. Puri, *On partially efficiency balanced designs II,* Commun. Statist.-Theor. Meth. 11(24), pp. 2817-2830, 1982.

Progresywnie modyfikowane statystyki porządkowe

Mariusz Bieniek

Instytut Matematyki
Uniwersytet Marii Curie-Skłodowskiej
Lublin

W referacie wprowadzony zostanie nowy model uporządkowanych danych statystycznych nazywany progresywnie modyfikowane statystyki porządkowe. Jest to rozszerzenie modelu progresywnie cenzurowanych statystyk porządkowych, w którym z próby usuwa się losowo wybrane elementy po uszkodzeniu jednego z nich. W modelu progresywnie modyfikowanym można w kolejnych krokach usuwać lub dodawać elementy do próby. Model ten służy jako ważny przykład uogólnionych statystyk porządkowych, gdyż podaje ogólną procedurę statystyczną pobierania danych empirycznych, której szczególnymi przypadkami są statystyki porządkowe i wartości rekordowe.

SLOPE - nowa procedura identyfikacji istotnych zmiennych w modelach liniowych

Małgorzata Bogdan

Instytut Matematyki i Informatyki
Politechniki Wrocławskiej

Załóżmy, że dysponujemy obserwacjami z modelu liniowego $y = X\beta + z$. SLOPE (Sorted L-One Penalized Estimation) proponuje identyfikację niezerowych elementów wektora β za pomocą minimalizacji wyrażenia

$$\frac{1}{2}\|y - Xb\|_{\ell_2}^2 + \lambda_1|b|_{(1)} + \lambda_2|b|_{(2)} + \ldots + \lambda_p|b|_{(p)},$$

gdzie $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_p$, a $|b|_{(1)} \geqslant |b|_{(2)} \geqslant \ldots \geqslant |b|_{(p)}$ są statystykami porządkowymi z wektora wartości bezwzględnych estymatorów współczynników regresji. Procedura ta jest podobna do popularnej procedury wielokrotnego testowania Benjaminiego-Hochberga. Zaprezentujemy wyniki teoretyczne gwarantujące kontrolę Frakcji Fałszywych Odkryć w przypadku gdy macierz eksperymentu jest ortogonalna oraz wyniki symulacyjne ilustrujące bardzo dobre własności tej metody w innych sytuacjach.

Prezentowane wyniki pochodzą z artykułu: M. Bogdan, E. van den Berg, W. Su and E.J. Candès, "Statistical Estimation and Testing via the Sorted l1 Norm".

Zastosowanie algorytmu filtru cząsteczkowego do zagadnień ekologii

Katarzyna Brzozowska - Rup

Katedra Matematyki
Politechniki Świętokrzyskiej

Antoni Leon Dawidowicz

Wydział Matematyki i Informatyki
Uniwersytetu Jagiellońskiego

Zagadnienia ekologiczne często są opisane za pomocą równań różniczkowych. Wiąże się to z dwoma problemami

1. W odróżnieniu od np. zjawisk fizycznych obserwacja przebiegu zjawiska jest daleko utrudniona.

2. Równania różniczkowe pojawiające się w modelu nie dają się efektywnie analitycznie rozwiązać, a algorytmy numeryczne są obarczone błędem.

Proponujemy następującą procedurę.

1. Co ustaloną jednostkę czasu wyznaczamy posiadany opis stanu układu.

2. Stan ten traktujemy jako zmienną obserwowaną.

3. Faktyczne rozwiązanie równania w danej chwili traktujemy, jako zmienną ukrytą.

Przykładem może być model "drapieżca - ofiara' w jego uogólnionej postaci. Wtedy zmienną obserwowaną można szacować w oparciu o dane z Kółek Łowieckich. Do estymacji parametrów strukturalnych modelu proponujemy zastosowanie sekwencyjnej metody Monte Carlo.

Literatura

[1] Katarzyna Brzozowska-Rup, Antoni Leon Dawidowicz, *Parameter Estimationfor Nonlinear State-Space Models Using Particle Method Combined with the EM Algorithm,* Financial Markets Principles of Modelling Forecasting and Decision-Making, 111-123, 2011

[2] A. Jasra, A. Doucet, *Sequential Monte Carlo methods for diffusion processes,* Proceedings of the Royal Society A 465,pp 3709-3727, 2009

[3] D. Rimmer, A. Doucet, W. J. Fitzgerald, *Particle Filters for Stochastic Differential Equations of Nonlinear Diffusions,* Technical Report, University of Cambridge, Engineering Dept, 2005

[4] Janusz Uchmański, *Klasyczna ekologia matematyczna,* PWN Warszawa 1992

Estymacja parametrów przełącznikowych modeli wielowymiarowych szeregów czasowych

Anna Czapkiewicz,
Paweł Jamer

AGH w Krakowie, Wydział Zarządzania

W zastosowaniach empirycznych często napotykamy rozkłady charakteryzujące się dużą kurtozą i silna asymetrią. Na przykład, do modelowania finansowych szeregów czasowych najczęściej używa się modele GARCH z warunkowym rozkładem skośnym t-Studenta. Konsekwencją tego faktu jest trudność konstruowania rozkładów wielowymiarowych, których składowe zachowują swoje własności. W literaturze przedmiotu, do szukania zależności pomiędzy danymi jednowymiarowymi procesami wykorzystuje się funkcje kopuli. W przypadku jednak finansowych wielowymiarowych szeregów czasowych dodatkowym utrudnieniem jest uwzględnienie dynamiki zmian takich procesów. W tym celu konstruuje się tzw. dynamiczne modele Copula-GARCH, w których owa dynamika sterowana jest przez ukryty proces Markowa. Problemem jest estymacja parametrów takiego modelu. Jednym z możliwych podejść jest zastosowanie metody największej wiarogodności z użyciem filtrów Hamiltona. W referacie zostanie przedstawiony sposób estymacji modelu dynamicznego z wykorzysta-niem filtrów Hamiltona. Zostaną rozważone wielowymiarowe kopule Gaussa z k liczbą reżimów procesu Markowa. Zostanie przedstawiony przykład empiryczny estymacji szeregów finansowych dla dwóch i dla trzech reżimów.

Literatura

[1] J. D. Hamilton, *Time Series Analysis,* Princeton University Press, 1994

[2] O. C. Filho, F.A. Ziegelmann, M.J.Dueker, *Modelling dependence dynamics through copulas with regime switching,* Insurance Mathematics and Economics,2012

Testowanie wielowymiarowej stochastycznej dominacji z wykorzystaniem wielowymiarowych funkcji kwantylowych

Kamil Dyba

Instytut Matematyczny Uniwersytetu Wrocławskiego

Niech dany będzie rozkład prawdopodobieństwa na $\mathbb{R}^d$ o dystrybuancie F_0 oraz próba $X_1, X_2, \ldots, X_n \in \mathbb{R}^d$ z rozkładu o dystrybuancie F. Rozważamy następujący problem testowania:

$$H : \forall x \in \mathbb{R}^d \; F(x) = F_0(x) \quad \text{vs} \quad K : \forall x \in \mathbb{R}^d \; F(x) \geqslant F_0(x) \; \wedge F \neq F_0.$$

W przypadku $d = 1$ znane są liczne rozwiązania powyższego zagadnienia. Jako przykład można tu podać jednostronny test Kołmogorowa-Smirnowa. Jego konstrukcja opiera się na tym, że przy hipotezie H rozkład statystyki $\sup_{x\in\mathbb{R}}(F_n(x) - F_0(x))$, gdzie F_n oznacza dystrybuantę empiryczną z próby $X_1, X_2, \ldots, X_n$, nie zależy od F (także w granicy przy $n \to \infty$).

W przypadku $d > 1$ powyższa własność nie zachodzi, co zmusza do poszukiwania rozwiązań tego problemu innych niż w przypadku jednowymiarowym. Rozwiązania dyskutowane w literaturze przedmiotu polegają na posługiwaniu się taką samą statystyką testową jak w przypadku $d = 1$ i wyznaczaniu jej rozkładu przy H w oparciu o metody symulacyjne.

W referacie zostanie zaproponowana konstrukcja testu alternatywna w stosunku do metod wykorzystujących symulacje, oparta o wielowymiarowe funkcje kwantylowe w ujęciu Einmahla i Masona.

Literatura

[1] Z. Guo, *Stochastic dominance and its applications in portfolio management,* 2012

[2] F. Hsieh, B. W. Turnbull, *Non- and semi-parametric estimation of the receiver operating characteristic curve,* Technical Report 1026, School of Opera- tions Research, Cornell University, 1992

[3] J. H. J. Einmahl, D. M. Mason, *Generalized quantile process,* Annals of Statistics, 20, pp. 1062-1078, 1992

Optymalne układy doświadczalne w modelach współoddziaływania

Katarzyna Filipiak,
Augustyn Markiewicz

Katedra Metod Matematycznych i Statystycznych
Uniwersytet Przyrodniczy w Poznaniu

W pracy przedstawiono charakterystyki układów optymalnych w modelach współoddziaływania; tzn. w modelach z efektami sąsiedztwa. Wykorzystano je następnie do zbadania optymalności układów zrównoważonych ze względu na wybrany schemat sąsiedztwa, których konstrukcje są przedmiotem licznych badań.

O pewnych własnościach układów wagowych

Małgorzata Graczyk

Uniwersytet Przyrodniczy w Poznaniu

W pracy przedstawiona zostanie tematyka estymacji nieznanych miar obiektów w modelu sprężynowego układu wagowego przy założeniu, że błędy pomiarów są nieskorelowane i mają różne wariancje.

Zostaną przedstawione i porównane warunki konieczne i dostateczne wyznaczające postaci macierzy układów optymalnych względem różnych kryteriów optymalności. Podsumowaniem rozważań jest podanie przykładowych metod konstrukcji macierzy układu optymalnego.

Literatura

[1] M. Graczyk, *Regular A-optimal spring balance weighing designs,* Revstat 10, pp.323-333, 2012

[2] K. Katulska, E. Przybył, *On certain D-optimal spring balance weighing designs,* Journal of Statistical Theory and Practice 1, pp.393-404, 2007

[3] K. Katulska, E. Rychlińska, *On regular E-optimality of spring balance weighing design,* Colloquium Biometricum 40, pp.165-176, 2007

Redukcja wymiarów danych za pomocą transformacji arctg

Jolanta Grala-Michalak

Wydział Matematyki i Informatyki UAM
Instytut Matematyki i Informatyki

W pracy opisano przekształcenie, które użyte jako krok wstępny w liniowej analizie dyskryminacji, powoduje albo redukcję wymiaru danych, albo poprawia wynik liniowej dyskryminacji w przypadku danych trudnoseparowalnych. Metoda jest prezentowana na dobrze znanych statystycznych zbiorach danych.

Literatura

[1] K. V. Mardia, P. Jupp, *Directional Statistics, 2 ed.,* Wiley, 2000

[2] J. Grala-Michalak, *Dimension reduction via arc tan transformation of data,* Comp. Stat. w przygotowaniu

Estymacja metodą największej wiarogodności w modelu liniowym z dwoma komponentami wariancyjnymi

Mariusz Grządziel

Katedra Matematyki Uniwersytetu Przyrodniczego we Wrocławiu

W normalnych modelach liniowych mieszanych z dwoma komponentami wariancyjnymi funkcja wiarogodności może mieć więcej niż jedno maksimum lokalne, więc zagadnienie wyznaczania wartości estymatorów największej wiarogodności w tych modelach należy do klasy problemów optymalizacji niewypukłej (por. [1] i [2]). W referacie zostanie pokazane, że zagadnienie to można sprowadzić do znajdowania wszystkich miejsc zerowych odpowiednio zdefiniowanego wielomianu.

Literatura

[1] E. Gross, M. Drton, S. Petrović, *Maximum likelihood degree of variance components models,* Electronic Journal of Statistics 6, s. 993-1016, 2012

[2] S. Welham, R. Thompson, *A note on bimodality in the log-likelihood function for penalized spline mixed models,* Computational Statistics & Data Analysis 53, s. 920-931, 2009

Testy zgodności dla nieprecyzyjnych danych

Przemysław Grzegorzewski

Wydział Matematyki i Nauk Informacyjnych
Politechniki Warszawskiej
oraz
Instytut Badań Systemowych PAN

Hubert Szymanowski

Interdyscyplinarne Studia Doktoranckie PAN

Niech $X_1, \ldots, X_n$ oznacza próbkę pochodzącą z nieznanego rozkładu F. Rozważmy problem testowania hipotezy $H : F = F_0$, gdzie F_0 jest pewną znaną dystrybuantą. Do weryfikacji tak postawionej hipotezy służą liczne testy zgodności, wśród których stosunkowo bogatą podklasę tworzą testy wykorzystujące dystrybuantę empiryczną.

Powszechnie stosowane testy zgodności zakładają, że dysponujemy precyzyjnymi pomiarami, tzn. że realizacje próby $x_1, \ldots, x_n$ są liczbami rzeczywistymi. Tymczasem w praktyce mamy często do czynienia z danymi nieprecyzyjnymi, których opis i analiza wymagają zastosowania aparatu wykraczającego poza ramy klasycznych metod statystyki.

W referacie zostanie przedstawiona konstrukcja testow zgodności bazujących na dystrybuanie empirycznej, pozwalających na weryfikację rozważanej hipotezy dotyczącej postaci rozkładu na podstawie nieprecyzyjnych danych, modelowanych za pomocą zbiorów rozmytych. W szczególności zostaną omówione uogólnienia kilku klasycznych testów zgodności, takich jak test Kolmogorowa, test Craméra-von Misesa oraz test Andersona-Darlinga.

Literatura

[1] P. Filzmoser, R. Viertl, *Testing hypotheses with fuzzy data. The fuzzy p-value,* Metrika 59, pp. 21-29, 2004

[2] P. Grzegorzewski, H. Szymanowski, *Goodness-of-fit tests for fuzzy data,* (zgłoszone do druku)

[3] G. Hesamian, S.M. Taheri, *Fuzzy empirical distribution function: properties and application,* Kybernetika (w druku)

Symulacyjne badanie testów wielowymiarowej normalności opartych na skośności i kurtozie

Zofia Hanusz

Katedra Zastosowań Matematyki i Informatyki, Uniwersytet Przyrodniczy w Lublinie

Rie Enomoto, Takashi Seo

Department of Mathematical Information Science, Tokyo University of Science

Kazuyuki Koizumi

Department of International College of Arts and Sciences, Yokohama City University

W pracy przedstawiono badania symulacyjne nad poziomem istotności i mocą testów do badania wielowymiarowej normalności opartych na skośności i kurtozie z próby. Wykorzystano testy oparte na statystykach Mardii i Srivastavy. Rozważono także test Jarque-Bera bazujący na mieszninie testów opartych na skośności i kurtozie. Rozważane w pracy testy porównano z testem Henze-Zirklera. Badania symulacyne przeprowadzono dla różnych liczebności prób i cech w obserwacjach dla poziomów istotności 0,05; 0,01 oraz 0,1.

Literatura

[1] Z. Hanusz, J. Tarasińska, Z. Osypiuk, *On the small sample properties of variants of Mardia's and Srivastava's kurtosis-based tests for multivariate normality,* Biometrical Letters 49(2), pp.159-175, 2012

[2] K. Koizumi, N.Okamoto, J., T. Seo, *On Jarque-Bera tests for assesing multivariate normality,* Journal of Statistics: Advances in Theory and Applications 1 (2), pp.207-220, 2009

Adaptacyjne testy wynikowe symetrii wokół nieznanego środka

Tadeusz Inglot,
Alicja Janic

Instytut Matematyki i Informatyki Politechniki Wrocławskiej

Problem testowania symetrii wokół nieznanego środka jest badany od dawna i uważany w literaturze za trudny. Większość istniejących testów jest opartych na porównaniu średniej i mediany z próby (np. Cabilio i Massaro, 1996, Zheng i Gastwirth, 2010) lub na pewnych uogólnieniach tej miary asymetrii (Ekström i Jammalamadaka, 2012). A więc wykrywa tylko pewne typy asymetrii. Zastosowanie teorii testów wynikowych pozwala skonstruować testy o szerokim spektrum czułości na różnego charakteru odstępstwa od symetrii.

W referacie przedstawimy konstrukcję statystyki wynikowej i odpowiednich reguł wyboru, dobór estymatorów parametrów zakłócających (w tym mediany), wyniki empiryczne, własności asymptotyczne oraz dyskusję zakresu stosowalności testów.

Adaptacyjne testy symetrii wokół znanego środka

Tadeusz Inglot,
Dawid Kujawa

Instytut Matematyki i Informatyki Politechniki Wrocławskiej

W pracy [1] zaproponowano i wszechstronnie przebadano adaptacyjne testy symetrii oparte na układzie wielomianów Legendre'a. Cechowały je stabilne, porównywalne z konkurencyjnymi testami moce dla typowych alternatyw oraz znacznie większe dla nietypowych alternatyw. Jednak dokładniejsza analiza charakteru asymetrii wielu alternatyw pokazuje, że wielomiany Legendre'a nie są najlepiej dopasowane do mierzenia asymetrii i sugeruje, że wybór odpowiedniego układu ortonormalnego pozwoli otrzymać lepszy test.

W referacie zaproponujemy układ ortonormalny funkcji, który realizuje powyższą sugestię. Przypomnimy konstrukcję statystyki wynikowej i reguł wyboru i omówimy własności asymptotyczne nowych testów. Przedstawimy wyniki badań symulacyjnych dla obszernej klasy alternatyw, zawierającej wszystkie alternatywy rozważane przez innych autorów oraz szereg nowych, reprezentujących różnorodne rodzaje asymetrii. Potwierdzają one w pełni oczekiwania. Nowe testy wykazują dużą czułość bez względu na rodzaj asymetrii, przez co można je zaliczyć do najlepszych testów typu omnibus. Przy okazji pokażemy, że test kierunkowy oparty na pierwszej funkcji rozważanego układu jest lepszy od testu Modarresa i Gastwirtha (1998) dla alternatyw o znacznej asymetrii na ogonach.

Literatura

[1] T. Inglot, A. Janic, J. Józefczyk, *Data driven tests for univariate symmetry,* Probab. Math. Statist. 32, pp.323-358, 2012

Charakteryzacje rozkładów za pomocą wybranych funkcji z teorii niezawodności

Maria Iwińska,
Magdalena Szymkowiak

Instytut Matematyki
Politechniki Poznańskiej

W pracy zostały przedstawione charakteryzacje rozkładów: wykładniczego, Weibulla oraz log-logistycznego. Charakteryzacje przeprowadzono za pomocą wartości oczekiwanych pewnych funkcji z teorii niezawodności.

Literatura

[1] S. Bhattacharjee, A.K. Nanda, S.Kr. Misra, *Inequalities involving expectations to characterize distributions,* Statistics and Probability Letters 83, pp.2113-2118, 2013

[2] A.K. Nanda, *Characterization of distribution through failure rate and mean residual life functions,* Statistics and Probability Letters 80, pp.752-755, 2010

Maksymalna wariancja k-tych rekordów

Krzysztof Jasiński

Wydział Matematyki i Informatyki
Uniwersytet Mikołaja Kopernika w Toruniu

Rozważamy próbę n niezależnych zmiennych losowych o tym samym rozkładzie, o ciągłej dystrybuancie i skończonej wariancji. Klimczak i Rychlik (2004) podali optymalne oszacowania dla wariancji zwykłych i k-tych rekordów w jednostkach wariancji pojedynczej zmiennej losowej. Poszukiwanie ich jest równoznaczne z poszukiwaniem maksimum pewnego wielomianu $f(x)$ na przedziale $(0,1)$. Nie udowodnili natomiast, że istnieje punkt $x_0 = x_0(k,n)$ taki, że wielomian rośnie na przedziale $(0,x_0)$, a maleje w $(x_0,1)$. To implikowałoby, że maksimum jest jedno i stanowi rozwiązanie równania $f^{'}(x) = 0$. Przedstawimy rozwiązanie tego problemu dla $1 \leqslant k \leqslant \max\{2, \frac{n^2+4n}{3n+4}\}$ oraz $n \geqslant 1$ stosując własność zmniejszania zmienności (*variation diminishing property*) funkcji potęgowych na przedziale $(0,+\infty)$. W tym celu rozszerzymy klasyczną własność zmniejszania zmienności skończonych kombinacji funkcji na przypadek nieskończony.

Literatura

[1] M. Klimczak, T. Rychlik, *Maximum variance of kth records,* Statist. Probab. Lett. 69, pp.421-430, 2004

[2] K. Jasiński, *Maximum variance of kth records,* submitted, 2013

Niezmienniczość porządku transformaty GTTT

Maria Kamińska-Zabierowska

Instytut Matematyczny
Uniwersytetu Wrocławskiego

Porządek uogólnionej transformaty TTT (GTTT - *generalized total time on test transform*) został wprowadzony przez Li i Shakeda, na wzór wcześniej rozważanych porządków, np. zwykłej transformaty TTT, jako punktowe porównanie dwóch transformat. Mówimy, że rozkład F jest mniejszy od rozkładu G w porządku transformaty GTTT indukowanej przez nieujemną funkcję h, jeśli dla wszystkich $u \in (0,1)$ zachodzi:

$$\int_{-\infty}^{F^{-1}(u)} h(F(x))dx \leqslant \int_{-\infty}^{G^{-1}(u)} h(G(x))dx,$$

o ile obie strony nierówności istnieją.

Przedstawimy operacje, względem których porządek transformaty GTTT jest niezmienniczy, m.in. wyniki wpisujące się w badania Bartoszewicza z Benduch i Skolimowską.

Literatura

[1] J. Bartoszewicz, M. Benduch, *Some properties of the generalized TTT transform,* J. Statist. Plann. Inference 139, pp. 2208-2217, 2009

[2] J. Bartoszewicz, M. Skolimowska, *Preservation of stochastic orders under mixtures of exponential distributions,* Probab. Engrg. Inform. Sci. 20, pp.655-666, 2006

[3] X. Li, M. Shaked, *A general family of univariate stochastic orders,* J. Statist. Plann. Inference 137, pp.3601-3610, 2007

[4] M. Shaked, J. Shanthikumar, *Stochastic orders,* Springer, New York, 2007

[5] M. Shaked, M. A. Sordo, A. Suárez-Llorens, *A class of location-independent variability orders, with applications,* J. Appl. Probab. Volume 47, pp.407-425, 2007

D-efektywność chemicznych układów wagowych przy skorelowanych błędach losowych

Krystyna Katulska,
Łukasz Smaga

Wydział Matematyki i Informatyki
Uniwersytet im. Adama Mickiewicza

W pracy rozważany jest problem estymacji nieznanych miar obiektów w chemicznym układzie wagowym w oparciu o kryterium D-optymalności. Przyjęto założenie, że błędy losowe są równo nieujemnie skorelowane i mają takie same wariancje. Udowodnione zostało twierdzenie określające górne ograniczenie wyznacznika macierzy informacji dla chemicznego układu wagowego. Warunki konieczne i dostateczne na to aby to ograniczenie było osiągnięte zostały również wykazane. Zaprezentowano także konstrukcję D-optymalnego układu wagowego, dla którego oszacowanie to jest osiągnięte. Nie dla wszystkich parametrów układu n i p takie układy istnieją. W takich sytuacjach zaproponowane zostały konstrukcje układów wagowych, które jak zostało pokazane charakteryzują się wysoką D-efektywnością. Prezentowane układy zostały porównane (pod względem D-efektywności) z układami uzyskanymi za pomocą znanego w literaturze algorytmu poszukiwania układów D-optymalnych a także wynikami poszukiwania układów D-optymalnych wśród dużej liczby układów generowanych losowo.

Literatura

[1] L. Angelis, E. Bora-Senta, C. Moyssiadis, *Optimal exact experimental designs with correlated errors through a simulated annealing algorithm,* Computational Statistics & Data Analysis 37, pp.275-296, 2001

[2] D.A. Bulutoglu, K.J. Ryan, *D-optimal and near D-optimal* 2^k *fractional factorial designs of resolution* V*,* Journal of Statistical Planning and Inference 139, pp.16-22, 2009

[3] K. Katulska, Ł. Smaga, *A note on D-optimal chemical balance weighing designs and their application,* Colloquium Biometricum 43, pp.37-45, 2013

[4] J. Masaro, C.S. Wong, *D-optimal designs for correlated random vectors,* Journal of Statistical Planning and Inference 138, pp.4093-4106, 2008

[5] M.G. Neubauer, R.G. Pace, *D-optimal* $(0, 1)$*-weighing designs for eight objects,* Linear Algebra and its Applications 432, pp.2634-2657, 2010

Analiza korelacji kanonicznych dla wielozmiennych danych funkcjonalnych

Mirosław Krzyśko,
Łukasz Waszak

Wydział Matematyki i Informatyki UAM w Poznaniu

W klasycznej analizie korelacji kononicznych obiekty są charakteryzowane za pomocą wielu cech obserwowanych w jednym punkcie czasowym. Chcemy wówczas znaleźć zależność pomiędzy wektorami $\boldsymbol{Y} \in \mathbb{R}^p$ oraz $\boldsymbol{X} \in \mathbb{R}^q$. W ostatnich latach rośnie zainteresowanie danymi reprezentowanymi za pomocą krzywych, tzw. danymi funkcjonalnymi (Ramsay i Silverman (2005)). W dotychczasowych pracach z tej tematyki obiekty są charakteryzowane za pomocą jednej cechy obserwowanej w wielu momentach czasowych, tj. interesuje nas zależność pomiędzy procesami losowymi $Y(t) \in L_2(I_1)$ oraz $X(t) \in L_2(I_2)$ (Krzyśko i Waszak (2013)).

W referacie zostanie zaproponowana konstrukcja zmiennych kanonicznych i korelacji kanonicznych dla wielozmiennych danych funkcjonalnych. Obiekty statystyczne będą charakteryzowane za pomocą wielu cech obserwowanych w wielu momentach czasowych (dane podwójnie wielowymiarowe). Dokładniej, zaprezentowana zostanie konstrukcja zmiennych kanonicznych i korelacji kanonicznych dla wielowymiarowych procesów stochastycznych $\boldsymbol{Y}(t) \in L_2(I_1)^p$ oraz $\boldsymbol{X}(t) \in L_2(I_2)^q$ oraz pokazany zostanie związek z klasyczną analizą kanoniczną.

Literatura

[1] Krzyśko, M., Waszak, Ł, *Canonical correlation analysis for functional data,* Biometrical Letters, 2013

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

Odporna estymacja parametrów w wielowymiarowym modelu normalnym - wyniki symulacji komputerowych

Agnieszka Kulawik

Instytut Matematyki
Uniwersytetu Śląskiego w Katowicach

W ramach referatu znajdą się wyniki symulacji komputerowych uzyskanych w celu porównania estymatora największej wiarogodności i odpornego estymatora wartości oczekiwanej i macierzy kowariancji wielowymiarowego rozkładu normalnego. Badania przeprowadzone zostały zarówno w przypadku danych modelowych, jak i po ich niewielkim zaburzeniu (5%). Omawiany odporny estymator został zbudowany w oparciu o zgodny w sensie Fishera i różniczkowalny w sensie Frécheta funkcjonał statystyczny.

Literatura

[1] T. Bednarski, S. Zontek, *Robust estimation of parameters in a mixed unbalanced model,* The Annals of Statistics 24 (4), pp. 1493-1510, 1996

[2] B.R. Clarke, *Uniqueness and Fréchet differentiability of functional solutions to maximum likelihood type equations,* The Annals of Statistics 11 (4), pp. 1196-1205, 1983

[3] A. Kulawik, S. Zontek, *Robust estimation in the multivariate normal model,* na ukończeniu

Wnioskowanie o dodatniej kwadrantowej zależnosci, I

Teresa Ledwina

Instytut Matematyczny PAN

Grzegorz Wyłupek

Instytut Matematyczny Uniwersytet Wrocławski

Przedstawimy konstrukcje dwóch nowych rozwiązań w problemie testowania hipotezy o istnieniu dodatniej kwadrantowej zależności przeciwko alternatywie orzekającej o jej braku. Zaprezentujemy uzyskane wyniki teoretyczne mówiące o kontroli błędu pierwszego rodzaju na całym zbiorze rozkładów z hipotezy przy ustalonej liczbie obserwacji oraz o zgodności jednego ze skonstruowanych testów statystycznych.

Literatura

[1] W. Albers, *Stop-loss premiums under dependence,* Insurance: Mathematics and Economics 24, pp. 173-185, 1999

[2] K. Behnen, *Asymptotic optimality and ARE of certain rank-order tests under contiguity,* Annals of Mathematical Statistics 42, pp. 325-329, 1971

[3] N. Blomqvist, *On a measure of dependence between two random variables,* Annals of Mathematical Statistics 21, pp. 593-600, 1950

[4] R. Cont, *Empirical properties of asset returns: stylized facts and statistical issues,* Quantitative Finance 1, pp. 223-236, 2001

[5] J. Dhaene, M. Denuit, S. Vanduffel, *Correlation order, merging and diversification,* Insurance: Mathematics and Economics 45, pp. 325-332, 2009

[6] J-D. Fermanian, D. Radulović, M. Wegkamp, *Weak convergence of empirical copula process,* Bernoulli 10, pp. 847-860, 2004

[7] I. Gijbels, M. Omelka, D. Sznajder, *Positive dependence tests for copulas,* Canadian Journal of Statistics 38, pp. 555-581, 2010

[8] T. Ledwina, G. Wyłupek, *Nonparametric tests for stochastic ordering,* TEST 21, pp. 730-756, 2012

[9] T. Ledwina, G. Wyłupek, *Validation of positive quadrant dependence,* przesłana do druku, 2013

[10] T. Yanagimoto, M. Okamoto, *Partial orderings of permutations and monotonicity of a rank correlation statistic,* Annals of the Institute of Statistical Mathematics 21, pp. 489-506, 1969

Wnioskowanie o dodatniej kwadrantowej zależności, II

Teresa Ledwina

Instytut Matematyczny PAN

Grzegorz Wyłupek

Instytut Matematyczny
Uniwersytet Wrocławski

Przedstawimy wyniki symulacji i zastosowanie do danych rzeczywistych nowych rozwiązań, omówionych w części I, i kilku innych, dostępnych w literaturze.

Zaprezentujemy również nową miarę kwadrantowej zależności i przedyskutujemy jej własności.

Literatura

[1] M. Denuit, O. Scaillet, *Nonparametric tests for positive quadrant dependence,* Journal of Financial Econometrics 2, pp. 422-450, 2004

[2] I. Gijbels, M. Omelka, D. Sznajder, *Positive dependence tests for copulas,* Canadian Journal of Statistics 38, pp. 555-581, 2010

[3] T. Ledwina, G. Wyłupek, *Validation of positive quadrant dependence,* przesłana do druku, 2013

[4] O. Scaillet, *A Kolmogorov-Smirnov type test for positive quadrant dependence,* Canadian Journal of Statistics 33, pp. 415-427, 2005

Wybór modelu liniowego poprzez jednoczesne usuwanie zmiennych ciągłych i łączenie poziomów zmiennych czynnikowych

Aleksandra Maj-Kańska, Agnieszka Prochenka

Instytut Podstaw Informatyki
Polskiej Akademii Nauk

Piotr Pokarowski

Wydział Matematyki, Informatyki i Mechaniki
Uniwersytetu Warszawskiego

Zagadnienie wyboru modelu jest zazwyczaj rozumiane jako selekcja zmiennych objaśniających, aczkolwiek gdy model zawiera zmienne czynnikowe pojęcie selekcji można rozumieć w dwojaki sposób: w celu zmnejszenia wymiarowości modelu można albo usunąć całą zmienną jakościową albo złączyć tylko niektóre jej poziomy.

W referacie zaprezentowany zostanie zachłanny algorytm DMR (Delete or Merge Regressors) pozwalający na jednoczesne usuwanie zmiennych ciągłych i łączenie poziomów faktorów w modelu. Algorytm ten jest uogólnieniem algorytmu zaproponowanego przez Zhenga i Loha (1995) i polega na hierarchicznej klasteryzacji hipotez liniowych postaci $\beta_{ik} = \beta_{jk}$ dla poziomów zmiennych czynnikowych i $\beta_i = 0$ dla zmiennych ciągłych, gdzie odległościami są kwadraty t-statystyk. Najlepszy model wybierany jest za pomocą kryterium informacyjnego.

DMR ma własność zgodności, a jego główną zaletą jest szybkość. Złożoność obliczeniowa algorytmu jest taka sama jak złożoność dopasowywania modelu liniowego i wynosi $O(np^2)$, gdzie n jest liczbą obserwacji, a p liczbą parametrów modelu. DMR jest kilkaset razy szybszy od innych algorytmów opisanych w literaturze (np. CAS-ANOVA, Bondell, Reich 2009).

Przedstawiony zostanie również pakiet środowiska `R` o nazwie `DMR`, w którym zaimplementowany jest algorytm DMR oraz jego wersja dla uogólnionych modeli liniowych.

Literatura

[1] H.D. Bondell, B.J. Reich, *Simultaneous factor selection and collapsing levels in anova,* Biometrics 65, 169-177. 34, 187-200, 2009

[2] X. Zheng, W.Y. Loh, *Consistent variable selection in linear models,* Journal of the American Statistical Association 90, 151-156, 1995

Estymacja parametru rozkładu statystycznego metodą maksymalizacji intensywności przetwarzania informacji

Marcin Makowski,
Edward W. Piotrowski

Instytut Matematyki
Uniwersytet w Białymstoku

Wydaje się, że jednym z priorytetów przyświecających efektywnemu gromadzeniu danych statystycznych jest pozyskanie informacji zgodne z zasadą „jak najwięcej, w jak najkrótszym czasie". Rozważamy proces przechwytywania jakichkolwiek danych porcjami. Przyjmijmy, że miara informacyjna tych porcji jest rzeczywistą zmienną losową o rozkładzie scharakteryzowanym parametrem, estymacją którego jesteśmy zainteresowani. Przyjmijmy też, że czynność przechwycenia bądź odrzucenia porcji (pakietu) danych zajmuje tą samą jednostkę czasu i także tyle samo czasu zabiera proces obróbki (czy przekazania dalej) przechwyconej porcji danych. W takim kontekście naturalną miarą intensywności przetwarzania informacji okazuje się następujący funkcjonał:

$$\rho(pdf_\gamma, a) = \frac{\int_a^\infty x \cdot pdf_\gamma(x+m)dx}{1+\int_a^\infty pdf_\gamma(x+m)dx},$$

gdzie: $pdf_\gamma(x)$ jest rozkładem gęstości prawdopodobieństwa zmiennej losowej będącej miarą porcji informacji, γ – parametrem rozkładu, m – pierwszym momentem rozkładu, a – parametrem ucięcia. Poszukiwanie maksimum tego funkcjonału ze względu na parametr ucięcia a prowadzi nas do zaskakującego twierdzenia o punkcie stałym:

> *Maksymalna wartość funkcji* $\rho(a) = \rho(pdf_\gamma, a)$ *leży w jej punkcie stałym. Taki punkt stały* a^* *istnieje i* $a^* > 0$.

Z równania na punkt stały wyznaczamy zależność parametru γ rozkładu $pdf_\gamma(x)$ jako funkcję wartości punktu stałego a^*.

W kontekście przyjętych założeń zagadnienie najskuteczniejszej estymacji parametru rozkładu jest niezachłane (non-greedy problem) – intensywność przetwarzania informacji jest większa gdy zamiast uwzględniać wszystkie pakiety informacyjne odrzucimy ich część tak, by parametrem obcięcia (odrzucenia danych) był punkt stały a^* intensywności $\rho(a)$. Ponieważ a^* przyciąga na całej dziedzinie funkcji $\rho(a)$, to naturalną techniką empirycznego

poszukiwania a^* jest iteracja $a_{k+1} = \rho(pdf_\gamma, a_k)$. Po znalezieniu empirycznej wartości a^*_{eff} z równania na punkt stały znajdujemy estymowaną wartość γ_{eff} parametru rozkładu $pdf_\gamma(x)$.

Stosowanie metody wyznaczenia parametru γ_{eff} przedstawione zostanie na przykładach dotyczących popularnych rozkładów: Gaussa, wykładniczego (i Pareto).

Literatura

[1] E. W. Piotrowski, *Natężenie zysku – model racjonalnego kupca,* Przegląd Statystyczny 46/2, pp.191-197, 1999.

[2] E. W. Piotrowski, J. Sładkowski, *A subjective supply-demand model: the maximum Boltzmann/Shannon entropy solution,* J. Stat. Mech., pp.1-16 (P03035), 2009.

O testowaniu hipotez liniowych w wielowymiarowych modelach liniowych

Andrzej Michalski

Katedra Matematyki Uniwersytetu Przyrodniczego we Wrocławiu

Autor rozważać będzie wielowymiarowy model liniowy postaci $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, gdzie $(n \times d)$ - wymiarowa macierz $\mathbf{Y}$ reprezentuje n obserwacji o d składowych, $\mathbf{X}$ jest $(n \times p)$ macierzą planu (w teorii eksperymentu) lub macierzą wartości zmiennych objaśniających (w analizie regresji), $\boldsymbol{\beta}$ jest $(p \times d)$ - macierzą nieznanych parametrów. Zakładamy, że $(n \times d)$- macierz $\boldsymbol{\epsilon}$ reprezentuje błędy losowe, ktore podlegają wielowymiarowemu rozkładowi normalnemu $\mathcal{N}_{n\times d}\{E(\boldsymbol{\epsilon}) = 0, Cov(\boldsymbol{\epsilon}) = I_n \otimes \boldsymbol{\Sigma}\}$, dla pewnej nieznanej dodatnio określonej $(d \times d)$- wymiarowej macierzy $\boldsymbol{\Sigma}$ ([1]). W referacie autor przedstawi rezultaty z zakresu testowania dowolnych hipotez liniowych dla parametrów dotyczących wartości oczekiwanej z aplikacjami([2]).

Literatura

[1] T. W. Anderson, *An introduction to multivariate statistical analysis,* 3rd edn., John Wiley & Sons, New York, 2003

[2] D. F. Morrison, *Wielowymiarowa Analiza Statystyczna,* PWN, Warszawa, 1990

Wokół twierdzenia Ruuda

Jan Mielniczuk

Instytut Podstaw Informatyki PAN
oraz
Wydział Matematyki i Nauk Informacyjnych
Politechniki Warszawskiej

Przyjmijmy, że model binarny

$$P(Y = 1|X) = q(\alpha + \beta^T X),$$

gdzie $q : R \to (0, 1)$ jest pewną funkcją odpowiedzi, $\beta \in R^p$ a $X \sim P_X$, został źle wyspecyfikowany. Załóżmy mianowicie przykładowo, że postulowaną funkcją odpowiedzi jest funkcja logitowa $p(s) = e^s(1 + e^s)^{-1}$ i $p \neq q$. Zagadnieniem, istotnym dla estymacji wektora β i selekcji istotnych zmiennych w modelu, jest kierunkowa zgodność w sensie Fishera metody największej wiarogodności dla β, sprowadzająca się do pytania, czy

$$(\alpha^*_{NW}, \beta^*_{NW}) := \text{argmin}_{(a,b)} E\{\Delta_X(q(\alpha + \beta^T X), p(a + b^T X))\},$$

gdzie $\Delta_X(q(\alpha + \beta^T X), p(a + b^T X)$ jest odległością Kullbacka-Leiblera rozkładów dwupunktowowych zadanych przez $P(Y = 1|X) = q(\alpha + \beta^T X)$ i $P(Y = 1|X) = p(a + b^T X)$, spełnia

$$\beta^*_{NW} = \gamma\beta \tag{2.1}$$

dla pewnego $\gamma \in R$.
Odpowiedź na to pytanie daje twierdzenie Ruuda, jednak przy istotnym warunku na rozkład P_X. Restrykcyjność warunku skłania do poszukiwania innych estymatorów β.

Omówione zostaną warunki (znalezione wspólnie z S. Jaroszewiczem i P. Teisseyre), przy których parametr oparty na maksymalizacji pola pod teoretyczną krzywą ROC ma własność (2.1).

Literatura

[1] Ruud, P., *Sufficient conditions for the consistency of maximum likelihood estimation despite misspecification of distribution in multinomial choice models,* Econometrica 1983, 51, 225-228

Oszacowania wariancji czasu życia systemów niezawodnościowych z permutowalnymi elementami

Patryk Miziuła

Uniwersytet Mikołaja Kopernika w Toruniu

Tomasz Rychlik

Instytut Matematyczny PAN
oraz
Uniwersytet Mikołaja Kopernika w Toruniu

Rozważamy systemy niezawodnościowe z permutowalnymi elementami. Założenie permutowalności elementów – znacznie słabsze od niezależności – pozwala na dobre modelowanie układów stosowanych w praktyce. Do tej pory oszacowania permutacji czasu życia znane były tylko dla systemów „k-spośród-n". W czasie referatu zostanie zaprezentowana metoda, której opracowanie umożliwiło osiągnięcie oszacowań dla ogólnych systemów. Oszacowania te są optymalne — graniczne wartości są osiągane przez powszechne w teorii niezawodności rozkłady potęgowe i Pareto. Uzyskane wyniki można uogólnić na inne miary rozproszenia, np. absolutne odchylenie od mediany czy miarę opartą na funkcji LINEX.

Literatura

[1] P. Miziuła, T. Rychlik, *Extreme dispersions of semicoherent and mixed system lifetimes,* submitted

[2] P. Miziuła, T. Rychlik, *Sharp bounds for lifetime variances of reliability systems with exchangeable components,* submitted

[3] T. Rychlik, *Extreme variances of order statistics in dependent samples,* Statist. Probab. Lett. 78, 1577–1582, 2008

Wklęsłe funkcje straty w kinetycznym modelowaniu reakcji chemicznych

Karol Opara

Instytut Badań Systemowych PAN

Pin Pin Oh

Faculty of Engineering
University of Nottingham, Malaysia Campus

Przebieg reakcji chemicznych w czasie opisuje się za pomocą modeli kinetycznych. Często wyrażają się one poprzez układy równań różniczkowych o znanej postaci. Identyfikacja ich parametrów wymaga rozwiązania zadania regresji nieliniowej. W tym celu powszechnie stosowana jest metoda najmniejszych kwadratów. Dane eksperymentalne często jednak obarczone są błędami losowymi o wyostrzonych, skośnych rozkładach oraz błędami systematycznymi. Powoduje to podatność metody najmniejszych kwadratów na przeuczenie, skutkujące nieadekwatnymi wielkościami wyznaczanych parametrów oraz niefizycznym przebiegiem krzywych będących rozwiązaniami modelu.

W pracy przedstawiono metodę identyfikacji parametrów równań kinetycznych wykorzystującą wklęsłe funkcje straty oraz regularyzację. Podejście takie charakteryzuje się zwiększoną odpornością na obserwacje obarczone znacznymi błędami. Zilustrowano je jakościowo na przykładzie danych eksperymentalnych przestawiających kinetykę ciągu reakcji chemicznych prowadzących do produkcji biodiesla z oleju palmowego. Ilościowe potwierdzenie skuteczności zaproponowanych metod oparto zaś o rozległe badania symulacyjne dla specjalnie przygotowanych zestawów problemów testowych.

Literatura

[1] K. Opara, P. P. Oh, *Kinetic and thermodynamic modelling of biodiesel reaction,* IIASA report, 1 Oct. 2012.

[2] C. Hennig, M. Kutlukaya, *Some thoughts about the design of loss functions,* REVSTAT–Statistical Journal, 5(1), pp. 19–39, 2007.

[3] H. Noureddini, D. Zhu, *Kinetics of transesterification of soybean oil,* Journal of the American Oil Chemists' Society, 74, pp. 1457–1463, 1997.

[4] P. T. Benavides, U. Diwekar, *Optimal control of biodiesel production in a batch reactor: Part I: Deterministic control,* Fuel 94, pp. 211–217, 2012.

Modelowanie wieku pacjentów za pomocą częstości metylacji cytozyny w DNA

Agnieszka Prochenka

Instytut Podstaw Informatyki
Polskiej Akademii Nauk

Metylacja to reakcja biochemiczna, która polega na przyłączaniu grup metylowych do cytozyny, rzadziej do adeniny w DNA. Metylacje cytozyny najczęściej zachodzą gdy cytozyna występuje obok guaniny na łańcuchu, tworząc tzw. dwunukleotydy CpG. Proces metylacji ma związek ze starzeniem się organizmu, a także występowaniem chorób takich jak nowotwory. Na zlecenie Policji, we współpracy z Zakładem Genetyki Medycznej Warszawskiego Uniwersytetu Medycznego prowadzone są badania nad modelem liniowym, który za pomocą niewielkiej liczby miejsc CpG (klikanaście) będzie modelował wiek osób na podstawie próbek krwi. Pozwoli to na obniżenie kosztów takiej analizy.

Referat będzie dotyczył wyników pracy nad tym problemem, do rozwiązania którego zostały użyte metody wyboru modelu liniowego przy $p >> n$ takie jak testy jednokrotne dla predyktorów, regularyzacja, czy metody krokowe. Porównanie skuteczności metod oparte zostało na minimalizacji estymatora błędu predykcji.

Literatura

[1] Hannum, Gregory, et al., *Genome-wide methylation profiles reveal quantitative views of human aging rates,* Molecular cell (2012).

[2] Ziller, Michael J., et al., *Charting a dynamic DNA methylation landscape of the human genome,* Nature 500.7463 (2013): 477-481.

Estymatory regresji rangowej oparte na metodzie LASSO

Wojciech Rejchel

Wydział Matematyki i Informatyki
Uniwersytet Mikołaja Kopernika w Toruniu

Regresja rangowa związana jest z zadaniami, w których należy przewidzieć lub rozpoznać uporządkowanie obiektów na podstawie obserwacji zmiennych pomocniczych, opisujących cechy tych obiektów. W pracy rozważamy modele, w których liczba cech może znacznie przerastać liczność próby. Estymatory regresji rangowej minimalizują ryzyko empiryczne z wypukłą funckją straty i, często używaną w problemach wielowymiarowych, karą typu LASSO [1]. Jakość predykcji procedur wyrażamy przy pomocy nierówności probabilistycznych dotyczących ich ryzyka [2].

Literatura

[1] R. Tibshirani, *Regression shrinkage and selection via the lasso,* Journal of the Royal Statistical Society: Series B 58, pp. 267-288, 1996

[2] S. van de Geer, *High-dimensional generalized linear models and the lasso,* Annals of Statistics 36, pp. 614-645, 2008

Rozpoznawanie sekwencji kodujących białko w genomach prokariotycznych za pomocą ukrytych łańcuchów Markowa

Piotr Sobczyk

Instytut Matematyki i Informatyki
Politechniki Wrocławskiej

W pracy przedstawiono sformalizowaną matematycznie, w stosunku do np. pracy Rabinera, teorię dotyczącą ukrytych łańcuchów Markowa oraz sposób ich wykorzystania do rozpoznawania sekwencji kodujących białko w genomach prokariotycznych. Wybrane algorytmy zostały zaimplementowane w języku C++. Ich jakość została porównana na rzeczywistych danych tj. genomie bakterii *Borrelia Burgdorferii*.

Literatura

[1] P. Sobczyk, *Ukryte łańcuchy Markowa i ich zastosowanie do wykrywania sekwencji kodujących białko w genomach prokariotycznych,* Praca magisterska, 2013

[2] L. R. Rabiner, *A tutorial on hidden markov models and selected applications in speech recognition,* Proceedings of the IEEE 77, no. 2, pp. 257–286, 1989

[3] M. Borodovsky, A. Lukashin, *Genemark.hmm: new solutions for gene finding,* Nucleic Acids Research 26, no. 4, 1107–1115, 1997

Diagonalna analiza dyskryminacyjna

Katarzyna Stąpor

Politechnika Śląska, Instytut Informatyki

Klasyczna analiza dyskryminacyjna (QDA/LDA) jest często wykorzystywana w budowaniu klasyfikatora dla danych wysokowymiarowych. Jednak w sytuacji, gdy wymiarowość przestrzeni cech d jest znacznie większa niż liczebność próby uczącej N ($d \gg N$), próbkowe macierze kowariancji są osobliwe, co uniemożliwia zastosowanie QDA/LDA. Dudoit [1] wprowadziła „diagonalne" wersje QDA/LDA zakładając niezależność cech, co jak się okazało ([1], [2]) prowadzi do efektywnego klasyfikatora. Nowa obserwacja X w diagonalnej LDA jest klasyfikowana do klasy i, która maksymalizuje wyrażenie:

$$\hat{g}_i^{DL}(X) = \sum_{j=1}^{d} (x_j - \hat{\mu}_{ij})^2 / \hat{\sigma}_j^2 - 2 \ln \hat{\pi}_i, \qquad i = 1, \ldots, c,$$

gdzie $\sum_i = \sum = \mathrm{diag}(\hat{\sigma}_1^2, \ldots, \hat{\sigma}_d^2)$ jest wspólną, próbkową macierzą kowariancji, $\hat{\pi}_i$, N_i — odpowiednio estymator prawdopodobieństwa a priori, liczba obserwacji w i-tej klasie, $\hat{\mu}_{ij}$ — j-tą składowa i-tej średniej próbkowej, c — liczbą klas. Można pokazać, że $\hat{g}_i^{DL}(X)$ jest estymatorem obciążonym, co w przypadku danych pochodzących z mikromacierzy DNA, gdzie stosunek d/N osiąga duże wartości może powodować znaczące błędy w predykcji. W komunikacie zostanie zaprezentowana wersja nieobciążona. Efektywność zaś nowej reguły zostanie zilustrowana wynikami badań symulacyjnych z użyciem odpowiednich estymatorów błędu predykcji (tj. dla danych niezrównoważonych, jakimi są dane mikromacierzowe).

Literatura

[1] S. Dudoit, J. Fridlyand, T. Speed, *Comparison of discrimination methods for the classification of tumors using gene expression data,* Journal of the American Statistical Association 97, pp. 77–87, 2002

[2] P.J. Bickel, E. Levina, *Some theory of Fisher's linear discriminant function, 'naive Bayes', and some alternatives when there are many more variables then observations,* Bernoulli 10, pp. 989–1010, 2004

[3] H. Haibo, E. Garcia, *Learning from imbalanced data,* IEEE Trans. Knowledge and data Engineering, 21, 9, pp. 1263–1284, 2009

Zmodyfikowane rozkłady prawdopodobieństwa typu j i ich charakterystyki

Katarzyna Steliga

Uniwersytet Marii Curie-Skłodowskiej w Lublinie

W referacie prezentujemy zmodyfikowane rozkłady typu j (dwumianowy, Poissona) oraz ich charakterystyki (momenty, współczynnik asymetrii i spłaszczenia). Rozważamy również ważone rozkłady wspomnianego typu i ich zastosowanie.

Literatura

[1] S. Berg, J. Jaworski, *Modified binomial and Poisson distributions with application in random mapping theory,* J. Statist. Plann. Infer. 18, pp. 313-322, 1988

[2] S. Berg, K. Nowicki, *Statistical inference for a class of modified power series distributions with applications to random mapping theory,* J. Statist. Plann. Infer. 28, pp. 247-261, 1991

[3] S. Chakraborty , *On Some New α-Modified Binomial and Poisson Distributions and Their Applications,* Comm. Statist. Theory Methods 37, pp. 1755-1769, 2008

[4] R. Gupta, *Modified power series distribution and some of its applications,* Sankhyā, Ser. B, 36, pp. 288–298, 1974

[5] M. Murat, D. Szynal, *Moments of discrete distributions via a differential operator,* Journal of Mathematical Sciences Vol. 191, No. 4, pp. 568-581, 2013

Estymacja sekwencyjna parametru położenia i potęg parametru skali

Agnieszka Stępień-Baran

Instytut Matematyki Stosowanej i Mechaniki
Uniwersytet Warszawski

W pracy przedstawiono problem estymacji sekwencyjnej parametru położenia dla rodzin z parametrem położenia oraz potęg parametru skali dla rodzin z parametrem skali w przypadku, gdy obserwacje dostępne są w chwilach losowych. Wyznaczono klasy optymalnych procedur sekwencyjnych przy założeniu, że strata związana z błędem estymacji jest określona przez zrównoważoną funkcję straty (balanced loss function) niezmienniczą odpowiednio względem parametru położenia i potęg parametru skali, a koszt obserwacji zdeterminowany jest przez funkcję momentu zatrzymania i liczbę obserwacji do tego momentu.

Uzyskano tzw. ściągające reguły zatrzymania (shrinkage stopping rules) poprawiające wyniki otrzymane w pracach Jokiel-Rokita & Stępień (2009) i Stępień-Baran (2009).

Literatura

[1] A. Jokiel-Rokita, A. Stępień, *Sequential estimation of a location parameter from delayed observations,* Statistical Papers 50, pp.363-372, 2009

[2] A. Stępień-Baran, *Sequential estimation of powers of a scale parameter from delayed observations,* Applicationes Mathematicae 36(1),pp. 13-28, 2009

Porównywanie układów blokowych ze względu na estymację kwadratową

Czesław Stępniak

Instytut Matematyki Uniwersytetu Rzeszowskiego

Zwykle uznaje się, że eksperyment liniowy z wektorem obserwacji X jest nie gorszy ze względu na estymację liniową niż eksperyment liniowy z wektorem obserwacji Y jeśli dla każdej formy liniowej wektora Y istnieje forma liniowa wektora X z taką samą wartością oczekiwaną i nie większą wariancją. Zakładając dodatkowo, że wektory obserwacji mają rozkłady normalne i zastępując formy liniowe przez formy kwadratowe otrzymujemy kryterium na to, aby jeden z eksperymentów był nie gorszy niż drugi ze względu na estymację kwadratową. Obydwa kryteria porównywania eksperymentów będą badane w kontekście układów blokowych.

Wykrywanie losowej liczby rozregulowań z zadana precyzją

Krzysztof Szajowski

Instytut Matematyki i Informatyki
Politechniki Wrocławskiej

Omawiane zagadnienie to sekwencyjne wykrywanie nieznanej, losowej, liczby jednorodnych stochastycznie sekwencji w ciągach zmiennych losowych zależnych. Jest to uogólnieni zagadnienia sformułowanego przez Kołmogorowa i opisanego w artykule Shiryaeva (1961) dla pewnych procesów. Przedstawione rozważania dotyczą takiego przypadku w którym kolejne obserwacje losowane są zgodnie z zadanym prawdopodobieństwem warunkowym, zależnym od aktualnie wylosowanego stanu. Mechanizm losowania jest przedziałami stały, a celem jest wykrycie i ocena momentów zmiany tego mechanizmu. Omówione zostaną wyniki opisane w notatce Ochman-Gozdek i Szajowski (2013).

Literatura

[1] G.V. Moustakides, *Quickest detection of abrupt changes for a class of random processes,* IEEE Trans. Inf. Theory 44, 1965–1968, 1998

[2] A. Ochman-Gozdek and K. Szajowski, *Detection of a random sequence of disorders,* Preprint I-18/2013, Instytut Matematyki i Informatyki, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, August 2013. 59th ISI World Statistics Congress 25–30 August 2013, Hong Kong Special Administrative Region, China

[3] W. Sarnowski and K. Szajowski, *On-line detection of a part of a sequence with unspecified distribution,* Stat. Probab. Lett., 78(15), 2511—2516, 2008

[4] A.N. Shiryaev, *The detection of spontaneous effects,* Sov. Math, Dokl., 2:740-743, 1961. translation from Dokl. Akad. Nauk SSSR 138, 799–801, 1961

Wybór punktów czasowych i układu blokowego w modelu krzywych wzrostu

Anna Szczepańska-Álvarez

Katedra Metod Matematycznych i Statystycznych,
Uniwersytet Przyrodniczy w Poznaniu

W pracy analizowane są dwa rodzaje układów: układ, będący zbiorem punktów pomiarowych wybranych z przedziału czasowego, zwanego dziedziną eksperymentu oraz układ blokowy związany z rozmieszczeniem obiektów w blokach. Celem pracy jest wyznaczenie tych układów w sposób optymalny. W związku z tym rozważane są kryteria A, D, E-optymalności. Ponadto prezentowane są relacje pomiędzy funkcjami informacji dla różnych układów.

Literatura

[1] A. Markiewicz, A. Szczepańska, *Optimal designs in multivariate linear models,* Statistics & Probability Letters 77, pp.426-430, 2007

[2] F. Pukelsheim , *Optimal Designs of Experiments,* Wiley, New York, 1993

[3] A. Szczepańska, *Simultaneous choice of time points and the block design in the growth curve model,* Statistical Papers 54(2), pp.413-425, 2013

Minimalizacja ryzyka empirycznego w problemach odwrotnych

Zbigniew Szkutnik

Wydział Matematyki Stosowanej AGH

Przedyskutujemy pochodzącą z teorii uczenia maszynowego ideę konstrukcji estymatorów przez minimalizację empirycznej wersji tzw. ryzyka, którego wersję dokładną minimalizuje estymowany obiekt. Okazuje się, że przybliżoną minimalizację wystarczy przeprowadzić na δ-sieciach i że często można w ten sposób otrzymać estymatory asymtotycznie minimaksowe i adaptacyjne. W przypadku problemów odwrotnych kluczowa jest odpowiednia konstrukcja funkcjonału ryzyka empirycznego. W tym kontekście omówimy niedawne wyniki Klemelä i Mammena, wskazując i poprawiając pewne błędy w oryginalnych twierdzeniach, a także wskażemy pewne zastosowania.

Literatura

[1] J. Klemelä, E. Mammen, *Empirical risk minimization in inverse problems,* Ann. Statist., 38, pp.482-511, 2010

[2] M. van der Laan, S. Dudoit, A.W. van der Vaart, *The cross-validated adaptive epsilon-net estimator,* Statistics and Decisions, 24, pp.373-395, 2006

Skąd się biorą „gorące" miejsca na genomie?

Piotr Szulc

Instytut Matematyki i Informatyki
Politechnika Wrocławska

W referacie przedstawię, czym są tak zwane „gorące" miejsca na genomie (ang. *hot-spots*), w jaki sposób mogą powstawać oraz co zrobić, by się ich pozbyć.

Literatura

[1] M. Bogdan, D.O. Siegmund, P. Szulc, H. Tang, R.W. Doerge, *Are "Hot-spots" an Indication of Complex Interactions?,* 2013

[2] P.M. Visscher, C.S. Haley, *Detection of quantitative trait loci in line crosses under infinitesimal genetic models,* Theor. Appl. Genet. 93, 691–702, 1996

Analiza skupień w detekcji genów warunkujących cechy ilościowe

Maciej Szydłowski

Katedra Genetyki i Podstaw Hodowli Zwierząt
Uniwersytet Przyrodniczy w Poznaniu

Marta Molińska-Glura

Katedra i Zakład Informatyki i Statystyki
Uniwersytet Medyczny w Poznaniu

Krzysztof Moliński

Katedra Metod Matematycznych i Statystycznych
Uniwersytet Przyrodniczy w Poznaniu

Przedstawiona zostanie eksploracja danych ilościowych związanych z lokalizacją tkanki tłuszczowej kilku ras świń. Grupowanie obiektów, do dalszej analizy, wykorzystuje metodę analizy skupień. Dla tak uzyskanych podzbiorów (profili), wykorzystując informacje molekularne, wykazano istotny wpływ struktury genetycznej na lokalizację tkanki tłuszczowej.

Literatura

[1] M. Switonski, M. Stachowiak, J. Cieslak, M. Bartz, M. Grzes, *Genetics of fat tissue accumulation in pigs – a comparative approach,* J Appl Genet 51:153–168, 2010

O normowanym estymatorze Dantziga

Hubert Szymanowski

Instytut Podstaw Informatyki PAN

Jedną z metod estymacji współczynników w modelu liniowym, stosowanych również w przypadku, gdy liczba obserwacji jest mniejsza od liczby zmiennych, jest estymator Dantziga dany wzorem:

$$\hat{\beta}_D = \arg\min_{\beta\in\mathbb{R}^p} \{|\beta| : \left|D^{-1}X^T(Y - X\beta)\right|_\infty \leqslant r\},$$

gdzie D jest macierzą diagonalną, zawierającą normy kolumn macierzy X, a r jest dodatnią stałą. W praktyce, przed obliczeniem powyższego estymatora, często normalizuje się kolumny macierzy eksperymentu X. Renormalizując otrzymane współczynniki, dostajemy estymator $\hat{\beta}_N$ dany wzorem:

$$D\hat{\beta}_N = \arg\min_{\beta\in\mathbb{R}^p} \{|D\beta| : \left|D^{-1}X^T(Y - X\beta)\right|_\infty \leqslant r\}.$$

Dla wielu metod estymacji, jak na przykład metody LASSO, uzyskany estymator nie ulega zmianie po normalizacji macierzy X i renormalizacji otrzymanych współczynników. W referacie omówiony zostanie przykład, pokazujący że w przypadku estymatorów $\hat{\beta}_D$ i $\hat{\beta}_N$ różnica między nimi może być dowolnie duża. Przedstawione zostaną także własności estymatora $\hat{\beta}_N$, znalezione wraz z Janem Mielniczukiem (por. [1]). Zostaną one porównane z własnościami estymatorów $\hat{\beta}_D$ oraz LASSO (por. [2]).

Literatura

[1] J. Mielniczuk, H. Szymanowski, *Some proporties of normalized Dantzig estimator,* w recenzji, 2013

[2] P. Bickel, Y. Ritov, A. Tsybakov, *Simultaneous analysis of Lasso and Dantzig selector,* Annals of Statistics, 37, pp.1705-1732, 2009

Wykorzystanie metod losowych podprzestrzeni do predykcji i selekcji zmiennych

Paweł Teisseyre

Instytut Podstaw Informatyki PAN

Metoda losowych podprzestrzeni (Random Subspace Method, w skrócie RSM) zaproponowana przez Ho w roku 1998 jest popularnym i efektywnym narzędziem klasyfikacji, szczególnie w sytuacji danych o wysokim wymiarze. W procedurze tworzy się komitet klasyfikatorów z których każdy jest zbudowany na losowo wybranym podzbiorze atrybutów. W pracy [2] wykorzystano metodę RSM do konstrukcji miary istotności zmiennych w wysoko-wymiarowym zadaniu regresji. W referacie przedstawię warianty oryginalnego podejścia, które pozwalają na wybór lepszych modeli. Zaprezentuję również wyniki symulacji w których porównuje działanie metod opartych na losowych podprzestrzeniach oraz inne popularne procedury selekcji zmiennych.

Wyniki uzyskano wspólnie z Janem Mielniczukiem.

Literatura

[1] T. K. Ho, *The random subspace method for constructing decision forests,* IEEE Trans. Pattern Anal. Machine Intell. 20 (8), pp.832-844, 1998

[2] J. Mielniczuk, P. Teisseyre, *Using random subspace method for prediction and variable importance assessment in linear regression,* Computational Statistics and Data Analysis, w druku, 2013

Najkrótsze przedziały ufności dla prawdopodobieństwa sukcesu w modelu ujemnym dwumianowym

Wojciech Zieliński

Katedra Ekonometrii i Statystyki
SGGW

W pracy przedstawiono konstrukcję najkrótszych przedziałów ufności dla prawdopodobieństwa sukcesu w modelu ujemnym dwumianowym.

Geometryczna zbieżność algorytmów Gibbsa

Iwona Żerda

Instytut Matematyki i Informatyki
Uniwersytetu Jagiellońskigo

Aktualnie dostępne są pewne teoretyczne wyniki dotyczące tempa zbieżności algorytmów Gibbsa, jednak w praktycznych zastosowaniach często weryfikacja nałożonych warunków jest bardzo trudna.

W pracy przeanalizowano tempo zbieżności tych algorytmów dla pewnej klasy rozkładów stacjonarnych rozważanych wcześniej w kontekście zbieżności algorytmu błądzenia losowego Metropolisa-Hastingsa. Przyjęte ograniczenie pozwaliło wskazać proste do sprawdzenia warunki wystarczające geometrycznej zbieżności algorytmów Gibbsa.

Zastosowanie sformułowanych stwierdzeń przedstawiono na przykładzie wybranych hierarchicznych modeli bayesowskich spotykanych w literaturze.

Literatura

[1] S. F. Jarner, E. Hansen, *Geometric ergodicity of Metropolis algorithms,* Stochastic Process. Appl. 85, pp. 341-361, 2000

[2] G. O. Roberts, R. L. Tweedie, *Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms,* Biometrika 83, pp. 95-110, 1996

[3] G. Fort , E. Moulines , G. O. Roberts , J. S. Rosenthal, *On the geometric ergodicity of hybrid samplers,* Journal of Applied Probability 40, pp. 123-146, 2003

Część III

Lista uczestników

1. **dr Mariusz Bieniek (str. 14)**
   Instytut Matematyki UMCS
   mariusz.bieniek@umcs.lublin.pl

2. **dr hab. Małgorzata Bogdan (str. 15)**
   Instytut Matematyki i Informatyki Politechniki Wrocławskiej
   malgorzata.bogdan@pwr.wroc.pl

3. **prof. Tomasz Burzykowski**
   Hasselt University
   tomasz.burzykowski@uhasselt.be

4. **prof. Bronisław Ceranka**
   Uniwersytet Przyrodniczy w Poznaniu
   bronicer@up.poznan.pl

5. **dr Anna Czapkiewicz (str. 18)**
   Wydział Zarzadzania Akademii Górniczo-Hutniczej
   im. Stanisława Staszica w Krakowie
   gzrembie@cyf-kr.edu.pl

6. **dr hab. Antoni Leon Dawidowicz (str. 16)**
   Uniwersytet Jagielloński
   antoni.leon.dawidowicz@im.uj.edu.pl

7. **mgr Kamil Dyba (str. 19)**
   Instytut Matematyczny Uniwersytetu Wrocławskiego
   kamil.dyba@math.uni.wroc.pl

8. **mgr Krzysztof Gachewicz**
   PZU Życie SA
   kgachewicz@pzu.pl

9. **dr Małgorzata Graczyk (str. 21)**
   Uniwersytet Przyrodniczy w Poznaniu
   magra@up.poznan.pl

10. **dr Jolanta Grala-Michalak (str. 22)**
    Uniwersytet im. Adama Mickiewicza w Poznaniu
    grala@amu.edu.pl

11. **dr Mariusz Grządziel (str. 23)**
Katedra Matematyki Uniwersytetu Przyrodniczego
we Wrocławiu
mariusz.grzadziel@up.wroc.pl

12. **prof. dr hab. Przemysław Grzegorzewski (str. 24)**
Wydział Matematyki i Nauk Informacyjnych
Politechniki Warszawskiej
pgrzeg@ibspan.waw.pl

13. **prof. dr hab. Zofia Hanusz (str. 25)**
Katedra Zastosowań Matematyki i Informatyki Uniwersytetu
Przyrodniczego w Lublinie
zofia.hanusz@up.lublin.pl

14. **prof. dr hab. Tadeusz Inglot (str. 26, 27)**
Instytut Matematyki i Informatyki Politechniki Wrocławskiej
tadeusz.inglot@pwr.wroc.pl

15. **dr Maria Iwińska (str. 28)**
Politechnika Poznańska
maria.iwinska@put.poznan.pl

16. **dr inż. Alicja Janic (str. 26)**
Instytut Matematyki i Informatyki Politechniki Wrocławskiej
alicja.janic@pwr.wroc.pl

17. **dr Krzysztof Jasiński (str. 29)**
Wydział Matematyki i Informatyki Uniwersytetu Mikołaja
Kopernika w Toruniu
krzys@mat.umk.pl

18. **mgr Maria Kamińska-Zabierowska (str. 30)**
Instytut Matematyczny Uniwersytetu Wrocławskiego
mazab@math.uni.wroc.pl

19. **dr Joanna Karłowska-Pik**
Instytut Podstaw Informatyki Polskiej Akademii Nauk
joanna.karlowska-pik@ipipan.waw.pl

20. **dr hab. Krystyna Katulska (str. 31)**
Wydział Matematyki i Informatyki Uniwersytetu
im. Adama Mickiewicza w Poznaniu
krakat@amu.edu.pl

21. **mgr Adam Kołacz**
Wydział Matematyki i Nauk Informacyjnych Politechniki Warszawskiej
ygmmasta@gazeta.pl

22. **prof. dr hab. inż. Jacek Koronacki**
Instytut Podstaw Informatyki Polskiej Akademii Nauk
jacek.koronacki@ipipan.waw.pl

23. **mgr Paweł Kozyra**
Instytut Matematyczny Polskiej Akademii Nauk
pawel_m_kozyra@wp.pl

24. **prof. dr hab. Mirosław Krzyśko (str. 32)**
Wydział Matematyki i Informatyki UAM w Poznaniu
mkrzysko@amu.edu.pl

25. **Dawid Kujawa (str. 27)**
Politechnika Wrocławska
daw.kujawa@gmail.com

26. **dr Agnieszka Kulawik (str. 33)**
Uniwersytet Śląski w Katowicach
agnieszka.kulawik@us.edu.pl

27. **prof. dr hab. Teresa Ledwina (str. 34, 36)**
Instytut Matematyczny Polskiej Akademii Nauk
ledwina@impan.pan.wroc.pl

28. **mgr Aleksandra Maj-Kańska (str. 37)**
Instytut Podstaw Informatyki Polskiej Akademii Nauk
a.maj@phd.ipipan.waw.pl

29. **dr Marcin Makowski (str. 39)**
Instytut Matematyki Uniwersytetu w Białymstoku
makowski.m@gmail.com

30. **prof. dr hab. Augustyn Markiewicz (str. 20)**
Uniwersytet Przyrodniczy w Poznaniu
amark@up.poznan.pl

31. **prof. dr hab. Iwona Mejza (str. 13)**
Uniwersytet Przyrodniczy w Poznaniu
imejza@up.poznan.pl

32. **prof. dr hab. Stanisław Mejza**
Uniwersytet Przyrodniczy w Poznaniu, Katedra Metod Matematycznych i Statystycznych
smejza@up.poznan.pl

33. **dr hab. Andrzej Michalski (str. 41)**
Katedra Matematyki Uniwersytetu Przyrodniczego we Wrocławiu
apm.mich@gmail.com

34. **prof. dr. hab. Jan Mielniczuk (str. 42)**
Instytut Podstaw Informatyki Polskiej Akademii Nauk
miel@ipipan.waw.pl

35. **mgr Patryk Miziuła (str. 43)**
Uniwersytet Mikołaja Kopernika w Toruniu
bua@mat.umk.pl

36. **mgr Przemysław Mogiłka**
PZU Życie SA
pmogilka@pzu.pl

37. **prof. dr hab. Krzysztof Moliński (str. 57)**
Katedra Metod Matematycznych i Statystycznych Uniwersytetu Przyrodniczego w Poznaniu
krys@up.poznan.pl

38. **mgr inż. Karol Opara (str. 44)**
Instytut Badań Systemowych Polskiej Akademii Nauk
karol.opara@ibspan.waw.pl

39. **mgr Agnieszka Prochenka (str. 37, 46)**
Instytut Podstaw Informatyki Polskiej Akademii Nauk
a.prochenka@phd.ipipan.waw.pl

40. **dr Wojciech Rejchel (str. 47)**
Wydział Matematyki i Informatyki Uniwersytetu Mikołaja Kopernika w Toruniu
iggypop@mat.umk.pl

41. **dr Magdalena Skolimowska-Kulig**
Instytut Matematyki i Informatyki Uniwersytetu Opolskiego
magdalena.skolimowska@math.uni.opole.pl

42. **dr Łukasz Smaga (str. 31)**
Wydział Matematyki i Informatyki Uniwersytetu
im. Adama Mickiewicza w Poznaniu
ls@amu.edu.pl

43. **mgr Piotr Sobczyk (str. 48)**
Instytut Matematyki i Informatyki Politechniki Wrocławskiej
piotr.sobczyk@pwr.wroc.pl

44. **mgr inż. Michał Sołtys**
Instytut Podstaw Informatyki Polskiej Akademii Nauk
msoltys@op.pl

45. **prof. dr hab. inż. Katarzyna Stąpor (str. 49)**
Instytut Informatyki Politechniki Śląskiej
katarzyna.stapor@polsl.pl

46. **mgr Katarzyna Steliga (str. 50)**
Wydział Ekonomiczny Uniwersytetu Marii Curie
Skłodowskiej w Lublinie
katarzyna.steliga@gmail.com

47. **dr inż. Agnieszka Stępień-Baran (str. 51)**
Instytut Matematyki Stosowanej i Mechaniki Uniwersytetu
Warszawskiego
a.stepien-baran@mimuw.edu.pl

48. **prof. dr hab. Czesław Stępniak (str. 52)**
Instytut Matematyki Uniwersytetu Rzeszowskiego
cees@ur.edu.pl

49. **dr hab. inż Krzysztof Szajowski (str. 53)**
Instytut Matematyki i Informatyki Politechniki Wrocławskiej
krzysztof.szajowski@pwr.wroc.pl

50. **dr Anna Szczepańska-Álvarez (str. 54)**
Katedra Metod Matematycznych i Statystycznych Uniwersytetu
Przyrodniczego w Poznaniu
sanna6@wp.pl

51. **dr hab. Zbigniew Szkutnik (str. 55)**
Akademia Górniczo-Hutnicza im. Stanisława Staszica
w Krakowie
szkutnik@agh.edu.pl

52. **mgr inż. Piotr Szulc (str. 56)**
Instytut Matematyki i Informatyki Politechniki Wrocławskiej
piotr.a.szulc@pwr.wroc.pl

53. **dr hab. Maciej Szydłowski (str. 57)**
Katedra Genetyki i Podstaw Hodowli Zwierząt Uniwersytetu Przyrodniczego w Poznaniu
mcszyd@jay.up.poznan.pl

54. **mgr Hubert Szymanowski (str. 24, 58)**
Instytut Podstaw Informatyki Polskiej Akademii Nauk
h.szymanowski@phd.ipipan.waw.pl

55. **dr Magdalena Szymkowiak (str. 28)**
Politechnika Poznańska
magdalena.szymkowiak@put.poznan.pl

56. **mgr Paweł Teisseyre (str. 59)**
Instytut Podstaw Informatyki Polskiej Akademii Nauk
teisseyrep@ipipan.waw.pl

57. **mgr Łukasz Waszak (str. 32)**
Uniwersytet im. Adama Mickiewicza w Poznaniu
lwaszak@amu.edu.pl

58. **dr Grzegorz Wyłupek (str. 34, 36)**
Instytut Matematyczny Uniwersytetu Wrocławskiego
wylupek@math.uni.wroc.pl

59. **prof. dr hab. Wojciech Zieliński (str. 60)**
Katedra Ekonometrii i Statystyki Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie
wojciech_zielinski@sggw.pl

60. **mgr Iwona Żerda (str. 61)**
Instytut Matematyki i Informatyki Uniwersytetu Jagiellońskiego
iwona.zerda@im.uj.edu.pl

Spis treści