XL Konferencja

STATYSTYKA MATEMATYCZNA Będlewo 2014

Będlewo, 1-5 grudnia 2014

Organizatorzy konferencji

Komisja Statystyki Matematycznej Komitetu Matematyki PAN
Międzynarodowe Centrum Matematyczne im. Stefana Banacha
Wydział Matematyki Stosowanej AGH

Komitet Naukowy

prof. dr hab. Teresa Ledwina - Przewodnicząca
prof. dr hab. Ryszard Magiera
prof. dr hab. Jan Mielniczuk

Komitet organizacyjny

dr Bogdan Ćmiel
dr Anna Dudek
dr Piotr Majerski
dr Konrad Nosek
dr hab. Zbigniew Szkutnik - Przewodniczący

Miejsce konferencji

Pałac Będlewo
Będlewo ul. Parkowa 1
62-060 Stęszew

Redakcja

Bogdan Ćmiel

Druk i oprawa

Tomatoe Print Shop

Część I

Przedmowa

Szanowni Państwo,

w roku bieżącym mamy przyjemność uczestniczyć w czterdziestej konferencji pod ogólną nazwą Statystyka Matematyczna. Konferencje z tego cyklu odbywały się w różnych miejscach w Polsce. Trzynaście z nich było dużymi międzynarodowymi spotkaniami. Więcej szczegółów znajdą Państwo na stronie WWW Komisji Statystyki Komitetu Matematyki PAN.

Taka okrągła liczba skłania do pewnych refleksji. Pierwszą z nich jest głęboka wdzięczność dla wszystkich, którzy z inicjatywą zorganizowania takiej konferencji ponad czterdzieści lat temu wystąpili i ją wdrożyli. Reprezentantem tej grupy w relacjach z Komitetem Matematyki PAN, który objął swym patronatem naszą konferencję, był prof. Tadeusz Caliński. Niestety, wielu organizatorów i uczestników pierwszych konferencji już nie ma wśród nas. Pamiętamy jednak o nich, o ich pracy i zasługach dla rozwoju statystyki w Polsce.

Jest rzeczą chyba dla wszystkich oczywistą, że Konferencja odegrała istotną rolę w budowaniu i rozwoju polskiego środowiska statystyki matematycznej oraz w jego konsolidacji. Równocześnie rosło w Polsce zapotrzebowanie na dobrze wykształconych statystyków. Tendencja ta jest nadal wyraźnie widoczna, co można zaobserwować w wynikach bieżących badań na temat polskiego rynku pracy (np. w ramach programu kierowanego przez prof. Elżbietę Kryńską). Taki trend może nas tylko cieszyć. Jest on również motywacją do dalszej pracy zarówno dla Komisji Statystyki Komitetu Matematyki PAN jak i nas wszystkich.

Organizatorami XL Konferencji Statystyka Matematyczna są:

- Komisja Statystyki Komitetu Matematyki PAN;
- Międzynarodowe Centrum Matematyczne im. Stefana Banacha;
- Wydział Matematyki Stosowanej AGH w Krakowie.

Dziękujemy im za pomoc i wsparcie w organizacji Konferencji.

Krajowa konferencja Statystyka Matematyczna odbywa się po raz pierwszy w Ośrodku Badawczo-Konferencyjnym IM PAN w Będlewie. Mamy nadzieję na to, że zagości ona w nim na dłużej.

Wszystkim Państwu życzymy owocnego uczestnictwa w Konferencji i miłego pobytu w Będlewie.

W imieniu Komitetu Naukowego i Komitetu Organizacyjnego

prof. Teresa Ledwina i dr hab. Zbigniew Szkutnik

Część II

Program konferencji

Poniedziałek, 1 grudnia 2014

08:00	**Śniadanie**
09:00	Otwarcie konferencji
09:15 - 10:00	**Tadeusz Bednarski** Statystyczna analiza przyczynowości I
10:00 - 10:30	**Przerwa**
10:30 - 10:50	**Wojciech Niemiro** Kawałkami deterministyczne procesy Markowa i algorytmy MCMC
10:50 - 11:10	**Błażej Miasojedow** Algorytm Metropolisa dla sieci bayesowskich z czasem ciągłym
11:10 - 11:30	**Kamil Dyba** Badanie wielowymiarowego porządku stochastycznego z wykorzystaniem pewnej charakteryzacji zbiorów górnych
11:30 - 11:50	**Przerwa**
11:50 - 12:10	**Bogdan Ćmiel**, Karol Dziedziul, Barbara Wolnik Test gładkości dla funkcji gęstości
12:10 - 12:30	Bogdan Ćmiel, **Karol Dziedziul**, Natalia Jarzębkowska Ortogonalny rozkład jedności na sferze i jego zastosowanie do estymacji gęstości
12:30 - 12:50	**Tomasz Górecki**, Mirosław Krzyśko, Waldemar Wołyński Wielowymiarowa funkcjonalna analiza regresji w zastosowaniu do klasyfikacji danych
13:00	**Obiad**
15:00 - 15:20	**Alicja Jokiel-Rokita**, Ryszard Magiera Rozkłady chwil zatrzymania w pewnych sekwencyjnych procedurach estymacji
15:20 - 15:40	Alicja Jokiel-Rokita, **Daniel Lazar**, Ryszard Magiera Predykcja bayesowska dla podwójnie stochastycznego procesu Poissona
15:40 - 16:00	**Krzysztof Szajowski** O agregacji indywidualnych decyzji w wielokryterialnych zadaniach optymalnego stopowania
16:00 - 16:30	**Przerwa**
16:30 - 17:15	**Tadeusz Bednarski** Statystyczna analiza przyczynowości II
18:00	**Kolacja**

Wtorek, 2 grudnia 2014

08:00	**Śniadanie**
09:00 - 09:45	**Tadeusz Bednarski** Statystyczna analiza przyczynowości III
09:45 - 10:30	**Przerwa**
10:30 - 10:50	**Paweł Marcin Kozyra**, Tomasz Rychlik KONKURS Ograniczenia na wartości oczekiwane L-statystyk wyrażone przez średnią różnicę Giniego
10:50 - 11:10	**Patryk Miziuła** KONKURS Oszacowania ilorazów rozproszeń mieszanek uporządkowanych rozkładów
11:10 - 11:30	**Hubert Szymanowski** KONKURS Zgodność selekcji Uogólnionego Kryterium Informacyjnego dla wysokowymiarowej regresji logistycznej
11:30 - 11:50	**Przerwa**
11:50 - 12:10	**Piotr Sobczyk** KONKURS Subspace clustering – nowa metoda iteracyjna
12:10 - 12:30	Alicja Jokiel-Rokita, **Rafał Topolnicki** KONKURS Semiparametryczna estymacja krzywej ROC
12:30 - 12:50	**Krzysztof Boczkowski**, Alicja Jokiel-Rokita, Monika Sikorska KONKURS Estymacja parametrów modulowanego procesu potęgowego
13:00	**Obiad**
15:00 - 15:20	Przemysław Grzegorzewski, **Adam Kołacz** KONKURS Miary rozproszenia dla danych wielowymiarowych
15:20 - 15:40	Alicja Szabelska, Idzi Siatkowski, **Joanna Zyprych-Walczak** KONKURS Przegląd problemów i metod związanych z analizą danych dotyczących ekspresji genów na podstawie eksperymentu głębokiego sekwencjonowania
15:40 - 16:00	**Michał Sołtys** KONKURS Algorytmy typu boosting w modelowaniu różnicowym
16:00 - 16:30	**Przerwa**
16:30 - 16:50	**Agnieszka Prochenka** KONKURS Modelowanie wieku pacjentów za pomocą częstości metylacji cytozyny w DNA - wyniki na podstawie danych z sekwencjonowania nowej generacji
17:00	**Posiedzenie Komisji Konkursowej**
18:00	**Kolacja**
19:00	**Posiedzenie Zespołu ds. Konferencji w Wiśle**

Środa, 3 grudnia 2014

08:00	**Śniadanie**
08:45 - 14:15	**Wycieczka**
14:30	**Obiad**
15:30 - 16:15	**Tadeusz Bednarski** Statystyczna analiza przyczynowości IV
16:15 - 16:45	**Przerwa**
16:45 - 17:05	**Małgorzata Bogdan** O granicach możliwości detekcji prawdziwego modelu za pomocą LASSO
17:05 - 17:25	**Wojciech Rejchel** Regresja rangowa z karą LASSO oraz nierówności z wyrocznią
17:25 - 17:45	**Joanna Karłowska-Pik** Klasyfikacja i selekcja zmiennych w regresji logistycznej metodą losowych podprzestrzeni
18:00	**Kolacja**

Czwartek, 4 grudnia 2014

08:00	**Śniadanie**
09:00 - 09:20	**Augustyn Markiewicz** Dostateczność statystyki kwadratowo dostatecznej w modelu mieszanym
09:20 - 09:40	**Katarzyna Filipiak** Testowanie hipotez o rozdzielnej strukturze macierzy wariancji-kowariancji w dwupoziomowym modelu wielowymiarowym
09:40 - 10:30	**Przerwa**
10:30 - 10:50	Mariusz Grządziel, **Andrzej Michalski** Estymacja komponentów wariancyjnych w liniowych modelach mieszanych metodą największej wiarogodności: istnienie ocen parametrów
10:50 - 11:10	**Mariusz Grządziel**, Andrzej Michalski Estymacja komponentów wariancyjnych w liniowych modelach mieszanych metodą największej wiarogodności: podejście algebraiczne
11:10 - 11:30	**Łukasz Smaga** Układy E-optymalne przy skorelowanych błędach losowych
11:30 - 11:50	**Przerwa**
11:50 - 12:10	**Roman Różański** Bootstrapowe predykcyjne obszary i przedziały ufności w modelach szeregów czasowych z zastosowaniem estymacji nieparametrycznej i semiparametrycznej
12:10 - 12:30	**Anna Dudek** Bootstrap blokowy dla współczynników funkcji autokowariancji okresowo skorelowanego szeregu czasowego
12:30 - 12:50	**Katarzyna Brzozowska-Rup**, Antoni Leon Dawidowicz Zastosowanie algorytmu filtru cząsteczkowego do prognozowania rozprzestrzeniania się zanieczyszczeń
13:00	**Obiad**
15:00 - 15:20	**Tomasz Rychlik** Optymalne oszacowania wartości oczekiwanych statystyk pozycyjnych z nieparametrycznych rodzin rozkładów generowanych przez porządek wypukły
15:20 - 15:40	**Agnieszka Goroncy** Oszacowania wartości oczekiwanych statystyk pozycyjnych i spacji w modelach z rosnącą intensywnością awarii
15:40 - 16:00	**Mariusz Bieniek** Oszacowania dla obciążenia wybranych L-statystyk
16:00 - 16:30	**Przerwa**
16:30 - 16:50	**Maria Iwińska**, Magdalena Szymkowiak Charakteryzacje rozkładów Weibulla
16:50 - 17:10	Maria Iwińska, **Magdalena Szymkowiak** Charakteryzacje odwróconych rozkładów Weibulla
17:10 - 17:30	**Krzysztof Jasiński** Oszacowania wariancji statystyk porządkowych pochodzących z modeli urnowych
18:30	**Uroczysta kolacja**

Piątek, 5 grudnia 2014

08:00	**Śniadanie**
09:00 - 09:20	**Teresa Ledwina**, Grzegorz Wyłupek O wielomianach Charliera w kontekście weryfikacji poissonowskości rozkładu obserwacji
09:20 - 09:40	**Grzegorz Wyłupek** Testowanie parasolkowatych alternatyw
09:40 - 10:00	**Karol Opara** Ocena dokładności współczynników modelu opartego na pomiarach stanu nawierzchni dróg
10:00 - 10:30	**Przerwa**
10:30 - 10:50	**Paweł Teisseyre** Wybrane metody klasyfikacji wieloetykietowej
10:50 - 11:10	**Piotr Szulc** Zgodność zmodyfikowanej wersji Bayesowkiego kryterium informacyjnego przy nienormalnym błędzie
11:10 - 11:30	**Zbigniew Szkutnik** O estymacji metodą minimalnego kontrastu w problemach odwrotnych.
11:30	Zakończenie konferencji
13:00	**Obiad**

Część III

Streszczenia

Wykład wiodący

Statystyczne metody analizy przyczynowości

Tadeusz Bednarski

Instytut Nauk Ekonomicznych, Uniwersytet Wrocławski

Od wielu lat obserwuje się rosnące zainteresowanie metodologią statystyczną związaną z weryfikacją przyczynowości. Dotyczy to zwłaszcza nauk społecznych i medycyny, gdzie nie ma możliwości pełnego planowania doświadczeń, a badacze muszą możliwie najefektywniej wykorzystać dane obserwacyjne. Celem wykładu jest zapoznanie słuchacza z tematyką przyczynowości w statystyce zarówno w ujęciu historycznym jak i w obszarze współcześnie rekomendowanych metod, gdzie wydaje się jest wiele wartościowych otwartych problemów badawczych. Przewiduje się omówienie następujących tematów:

1. Filozofia przyczynowości
2. Randomizacja i przyczynowość (J. Neyman, R. Fisher)
3. Pojęcie przyczynowości w statystyce (P. Suppes, J. Pearl, D. Cox)
4. Współczesne metody analizy przyczynowości (D. Rubin, O. Aalen, T. Schweder, C. Granger)
5. Przykład badań statystycznych związanych z weryfikacją mechanizmu odmowy w badaniach ankietowych.

Referaty

Oszacowania dla obciążenia wybranych L–statystyk

Mariusz Bieniek

Instytut Matematyki, Uniwersytet Marii Curie-Skłodowskiej w Lublinie

Niech $X_{1:n} \leqslant \ldots \leqslant X_{n:n}$ oznaczają statystki porządkowe losowej próby prostej rozmiaru n z rozkładu F o średniej μ i wariancji σ^2. Rozważamy problem estymacji nieznanej średniej μ przy użyciu L–statystyk, czyli kombinacji liniowych statystyk porządkowych. W szczególności, dla ustalonych $1 \leqslant r \leqslant s \leqslant n$ określamy: średnią winsorowaną

$$W_{r,s:n} = \frac{1}{n}\left(rX_{r+1:n} + \sum_{i=r+1}^{s} X_{i:n} + (n-s)X_{s:n}\right)$$

quasi–średni rozstęp z próby (*sample quasimidrange*)

$$M_{r,s,:n} = \frac{1}{2}\left(X_{r:n} + X_{s:n}\right),$$

oraz trójśrednią próby (*sample trimean*)

$$T_n = \frac{1}{4}\left(X_{\lfloor \frac{n}{4} \rfloor + 1:n} + 2M + X_{n-\lfloor \frac{n}{4} \rfloor :n}\right),$$

gdzie M oznacza medianę próby. W referacie przedstawimy optymalne oszacowania dla obciążenia estymacji średniej μ przez powyższe L–statystyki. Oszacowania te zostały uzyskane przy użyciu metody projekcji. Omówimy również wnioski płynące z numerycznie wyznaczonych wartości uzyskanych oszacowań.

Estymacja parametrów modulowanego procesu potęgowego

Krzysztof Boczkowski, Alicja Jokiel-Rokita, Monika Sikorska

Wydział Podstawowych Problemów Techniki, Politechnika Wrocławska

Załóżmy, że zdarzenia (np. awarie) występują zgodnie z procesem Weibulla, załóżmy ponadto, że obserwujemy tylko co k-te zdarzenie, gdzie k jest dodatnią liczbą całkowitą. Obserwowany ciąg $\{T_i, i = 1, 2, \dots\}$ czasów $T_1, T_2, \dots$ nazywamy modulowanym procesem potęgowym (MPLP). Proces ten został zaproponowany przez Lakey'a i Rigdona (1992) jako kompromis pomiędzy niejednorodnym procesem Poissona a procesem odnowy. Stanowi on szczególny przypadek procesu odnowy z trendem zaproponowanego przez Lindqvista i innych (2003).

Problem estymacji nieznanych parametrów procesu MPLP metodą największej wiarogodności rozpatrywany był między innymi przez Lakey'a i Ringdona (1992). Kolejna metoda estymacji, metoda estymatora prostego, została przedstawiona przez Bandyopadhyay'a i Sena (2005).

W referacie przedstawimy alternatywną metodę estymacji nieznanych parametrów oparta na uogólnionej metodzie momentów (GMM). Zaprezentujemy również wyniki symulacyjnego porównania rozważanych estymatorów parametrów procesu MPLP oraz estymatorów wartości oczekiwanej czasu następnego skoku tego procesu.

Literatura

[1] N. Bandyopadhyay, A. Sen, *Non-standard asymptotic in an inhomogeneous gamma process,* Ann. Inst. Statist. Math., 57(4):703-732, 2005.

[2] S. E. Black, S. E. Rigdon, *Statistical inference for modulated power law process,* Journal of Quality Technology, 28(1):81-90, 1996.

[3] A. Jokiel-Rokita, R. Magiera, *Estimation of parameters for trend-renewal processes,* Statistics and Computing, 22:625-637, 2012.

[4] M. J. Lakey, S. E. Rigdon, *The modulated power law process,* Procedings of the 45th Annual Quality Congress, 559-563, 1992.

[5] B.H. Lindqvist, G. Elvebakk, K. Heggland, *The trend-renewal process for statistical analysis of repairable systems,* Technometrics, 45(1):31-44, 2003.

O granicach możliwości detekcji prawdziwego modelu za pomocą LASSO

Małgorzata Bogdan

Wydział Podstawowych Problemów Techniki, Politechnika Wrocławska

Przedstawimy asymptotyczne wyniki teoretyczne umożliwiające predykcję własności LASSO w sytuacji gdy elementy macierz eksperymentu są niezależnymi zmiennymi losowymi o rozkładzie gaussowskim. W szczególności zademonstrujemy krzywą opisującą tzw. minimaksową frakcję fałszywych odkryć (False Discovery Rate) jako funkcję parametru rzadkości. Minimaksowe FDR odpowiada rozważaniu najgorszego możliwego scenariusza ze względu na rozkład wielkości sygnałów dla ustalonej wartości parametru wygładzającego λ i następnie optymalizację tego najgorszego $FDR(\lambda)$ ze względu na λ. Przedyskutujemy również dolne ograniczenie na ważoną sumę błędów pierwszego i drugiego rodzaju w sytuacji gdy λ jest wybierane adaptacyjnie ze względu na wielkość sygnału.
Wyniki te zostały uzyskane we współpracy z Weijie Su i Emmanuelem Candésem z Uniwersytetu Stanforda.

Zastosowanie algorytmu filtru cząsteczkowego do prognozowania rozprzestrzeniania sę zanieczyszczeń

Katarzyna Brzozowska - Rup

Wydział Zarządzania i Modelowania Komputerowego, Politechnika Świętokrzyska

Antoni Leon Dawidowicz

Wydział Matematyki i Informatyki, Uniwersytet Jagielloński

W pracy [3] przedstawiono algorytm prognozowania zanieczyszczeń atmosfery. Oparty jest on o równanie adwekcji - dyfuzji. Ponieważ jest on rozwiązywany za pomocą algorytmu numerycznego błąd algorytmu można potraktować, jako zmienną losową i, co za tym idzie sam proces, jako proces Markowa. Ponieważ zanieczyszczenia są monitorowane sam poziom zanieczyszczeń można traktować, jako zmienne ukryte, a wyniki monitoringu jako zmienne obserwowane. Celem naszego referatu będzie zaproponowanie zastosowania do tego modelu algorytmu filtru cząsteczkowego opisanego m. in. w pracach [1,2]

Literatura

[1] K. Brzozowska-Rup, A. L. Dawidowicz, *Metoda filtru cząsteczkowego,* Matematyka Stosowana: matematyka dla społeczeństwa vol. 10(51) (2009), 69-107

[2] K. Brzozowska-Rup, A. L. Dawidowicz, *Parameter Estimationfor Nonlinear State-Space Models Using Particle Method Combined with the EM Algorithm,* Financial Markets Principles of Modelling Forecasting and Decision-Making (2011), 111-123

[3] K. Twardowska, I. Rożek, *Teoria procesów stochastyczntch w prognozowaniu zanieczyszczenia atmosfery,* Prz. Geof. XXXIII, 2 (1988), 139 - 151

Badania finansowane w ramach grantu NCN nr UMO-2013/11/D/HS4/04014.

Ortogonalny rozkład jedności na sferze i jego zastosowanie do estymacji gęstości

Bogdan Ćmiel, Karol Dziedziul, Natalia Jarzębkowska

Wydział Matematyki Stosowanej, Akademia Górniczo-Hutnicza w Krakowie,
Wydział Fizyki Technicznej i Matematyki Stosowanej, Politechnika Gdańska

Estymacja na sferze oraz szerzej na rozmaitości była w ostanim okresie bardzo intensywnie badana ze względu na rozwój framek (ang. frames) zwanych needletami, patrz literatura. W ostatnim czasie pojawiły się konkurencyjne framki powstałe przy użyciu ortogonalnego rozkładu jedności. Chciałbym przystępnie opowiedzieć o tym czym są framki oraz jak wyglada konstrukcja framek na sferze. Na końcu zasygnalizować ich zastosowanie do konstrukcji adaptacyjnego estymatora funkcji gęstości oraz analizy parametru gładkości.

Literatura

[1] M. Bownik, K.Dziedziul, *Smooth orthogonal projections on sphere,* ukaże się w Constructive Approximation

[2] A.D. Bull, R. Nickl, *Adaptive confidence sets in L2,* Probab. Theory Related Fields 156 (3–4) (2013) 889–919

[3] G. Kerkyacharian, R. Nickl, D. Picard, *Concentration inequalities and confidence bands for needlet density estimators on compact homogeneous manifolds,* Probab. Theory Related Fields 153 (1–2) (2012) 363–404.

[4] A. Kueh, *Locally Adaptive Density Estimation on the Unit Sphere Using Needlets,* Constr. Approx., 36 (2012), no. 3, 433-458.

Test gładkości dla funkcji gęstości

Bogdan Ćmiel, Karol Dziedziul, Barbara Wolnik

Wydział Matematyki Stosowanej, Akademia Górniczo-Hutnicza w Krakowie,
Wydział Fizyki Technicznej i Matematyki Stosowanej, Politechnika Gdańska,
Instytut Matematyki, Uniwersytet Gdański

Rozważmy próbę prostą z rozkładu zadanego funkcją gęstości z przestrzeni L^2. Przedstawiony zostanie problem testowania hipotezy H_0: funkcja gęstości ma nie więcej niż μ słabych pochodnych, przeciwko alternatywie H_1: funkcja gęstości ma więcej niż μ słabych pochodnych. Do konstrukcji testu zostanie użyty asymptotyczny rozkład estymatora parametru gładkości. Aby uzyskać ten rozkład zaproponowana zostanie pewna metoda wzbogacania próby. Działanie testu zostanie zademonstrowane w eksperymencie numerycznym.

Literatura

[1] B. Ćmiel, K. Dziedziul, B. Wolnik, *The smoothness test for a density function,* Nonlinear Analysis: Theory, Methods & Applications, 104, 21-39, 2014.

[2] K. Dziedziul and B Ćmiel, *Density smoothness estimation problem using a wavelet approach,* ESAIM Probability and Statistics, 18, 130 - 144, 2014.

[3] K. Dziedziul, M. Kucharska, B. Wolnik, *Estimation of the smoothness of density,* J. Nonparametr. Stat., 23 (4), 991-1001, 2011.

[4] A.D. Bull, R. Nickl, *Adaptive confidence sets in L^2,* Probab. Theory Related Fields, 156 (3-4), 889-919, 2013.

Bootstrap blokowy dla współczynników funkcji autokowariancji okresowo skorelowanego szeregu czasowego

Anna Dudek

Wydział Matematyki Stosowanej, Akademia Górniczo-Hutnicza w Krakowie

Przedstawiony zostanie problem estymacji współczynnków funkcji autokowariancji okresowo skorelowanego szeregu czasowego. Omówione zostanie zastosowanie bootstrapu blokowego do konstrukcji punktowych i jednoczesnych przedziałów ufności dla tych współczynników. W szczególności zostaną przedstawione wyniki zgodności dla uogólnionego bootstrapu bloków sezonowych (Generalized Seasonal Block Bootstrap) oraz bootstrapu bloków ruchomych (Moving Block Bootstrap). Prezentacja będzie wzbogacona o wyniki przeprowadzonych badań symulacyjnych.

Literatura

[1] A.E. Dudek, *Circular block bootstrap for coefficients of autocovariance function of almost periodically correlated time series,* Metrika,
DOI 10.1007/s00184-014-0505-9, 2014.

[2] A.E. Dudek, S. Maiz, M. Elbadaoui, *Generalized Seasonal Block Bootstrap in frequency analysis of cyclostationary signals,* Signal Processing, 104C, 358-368, 2014.

[3] A.E. Dudek, J. Leśkow, D. Politis, E. Paparoditis, *A generalized block bootstrap for seasonal time series,* Journal of Time Series Analysis, 35, 89-114, 2014.

Badanie wielowymiarowego porządku stochastycznego z wykorzystaniem pewnej charakteryzacji zbiorów górnych

Kamil Dyba

Instytut Matematyczny, Uniwersytet Wrocławski

Wielowymiarowy porządek stochastyczny ($\leqslant_{st}$) najczęściej bywa definiowany w literaturze w następujący sposób: dla dwóch wektorów losowych X i Y o wartościach w $\mathbb{R}^d$ mówimy, że $X \leqslant_{st} Y$, gdy $E\phi(X) \leqslant E\phi(Y)$ dla dowolnej niemalejącej po współrzędnych borelowskiej funkcji $\phi\colon \mathbb{R}^d \to \mathbb{R}$ takiej, że powyższe wartości oczekiwane istnieją.

Wprost z definicji wynika, że porządek ten zachowuje się przy przekształcaniu wektorów losowych przez funkcje niemalejące, a i sama własność definiująca go jest pożądana przy modelowaniu wielu zjawisk np. w teorii ubezpieczeń. Niestety wykazanie uporządkowania dwóch rozkładów w tym porządku w oparciu o powyższą definicję nastręcza trudności. Z tego powodu często zamiast wielowymiarowego porządku stochastycznego rozważa się silniejszy od niego *wielowymiarowy porządek ilorazu wiarogodności*.

Znana jest również następująca charakteryzacja wielowymiarowego porządku stochastycznego: $X \leqslant_{st} Y$ wtedy i tylko wtedy, gdy $P(X \in U) \leqslant P(Y \in U)$ dla każdego zbioru górnego U tzn. borelowskiego podzbioru $\mathbb{R}^d$ o tej własności, że jeśli $u \in U$ i $u \leqslant v$ (w porządku po współrzędnych), to $v \in U$. Aby efektywnie móc wykorzystać tę własność, należy scharakteryzować zbiory górne w $\mathbb{R}^d$.

W referacie zostanie zaprezentowana charakteryzacja domkniętych zbiorów górnych w $\mathbb{R}^d$ wraz z jej praktycznym zastosowaniem w badaniu wielowymiarowego porządku stochastycznego.

Literatura

[1] C. D. Aliprantis, K. C. Border, *Infinite Dimensional Analysis: A Hitchhiker's Guide,* Springer Science & Business Media, 2007.

[2] M. Shaked, J. Shantikumar, *Stochastic Orders,* Springer, 2007.

Testowanie hipotez o rozdzielnej strukturze macierzy wariancji-kowariancji w dwupoziomowym modelu wielowymiarowym

Katarzyna Filipiak

Katedra Metod Matematycznych i Statystycznych,
Uniwersytet Przyrodniczy w Poznaniu

W zagadnieniu testowania hipotez o rozdzielnej strukturze macierzy wariancji-kowariancji macierzy obserwacji dla danych dwupoziomowych w literaturze używa się zwykle testu ilorazu wiarogodności (LRT). Celem mojego referatu jest przedstawienie zastosowania testu „Rao score" (RST) do testowania hipotezy o rozdzielnej strukturze macierzy wariancji-kowariancji, przy czym jej pierwszy czynnik ma postać kompletnie symetrycznej macierzy korelacyjnej, przy założeniu wielowymiarowego rozkładu normalnego obserwacji. Pokazana zostanie m.in. postać statystyki RST, a także za pomocą badań symulacyjnych porównane zostaną empiryczne rozkłady statystyk testowych LRT i RST oraz ich empiryczne prawdopodobieństwa błędu I-go rodzaju, w zależności od liczebności próby. Ponadto obydwa testy zilustrowane zostaną rzeczywistymi danymi.

Literatura

[1] K. Filipiak, D. Klein, A. Roy, *Score test for a separable covariance structure with the first component as compound symmetric correlation matrix,* Przesłana do redakcji.

Oszacowania wartości oczekiwanych statystyk pozycyjnych i spacji w modelach z rosnącą intensywnością awarii

Agnieszka Goroncy

Wydział Matematyki i Informatyki, Uniwersytet Mikołaja Kopernika w Toruniu

Niech $X_1, \ldots, X_n$ będą niezależnymi zmiennymi losowymi o tym samym rozkładzie z dystrybuantą F, która poprzedza dystrybuantę standardowego rozkładu wykładniczego W w porządku wypukłym. Oznacza to, że F ma rosnącą intensywność awarii (IFR). W pracy prezentujemy optymalne górne oszacowania wartości oczekiwanych pojedynczych statystyk pozycyjnych oraz spacji opartych na ciągu $X_1, \ldots, X_n$, wyrażone w terminach średniej z populacji oraz odchylenia standardowego.

Powyższe wyniki uzyskano we współpracy z prof. Tomaszem Rychlikiem (IM PAN).

Literatura

[1] Goroncy, A., Rychlik, T., *Evaluations of expected order statistics and spacings based on the IFR distributions,* w przygotowaniu.

[2] Goroncy, A., Rychlik, T., *Optimal bounds on expectations of order statistics and spacings from nonparametric families of distributions generated by convex transform order,* Metrika, DOI:10.1007/s00184-014-0495-7, 2014.

Wielowymiarowa funkcjonalna analiza regresji w zastosowaniu do klasyfikacji danych

Tomasz Górecki, Mirosław Krzyśko, Waldemar Wołyński

Wydział Matematyki i Informatyki,
Uniwersytet im. Adama Mickiewicza w Poznaniu

Wielowymiarowymi danymi funkcjonalnymi nazywamy dane w postaci wektora funkcji ciągłych określonych na danym przedziale. Dane tego typu są traktowane jako realizacje wielowymiarowego procesu losowego. W pracy przedstawiona zostanie nowa metoda klasyfikacji wykorzystująca techniki wielokrotnej analizy regresji (liniowa, logistyczna, nieparametryczna) dla tego typu danych. W przypadku więcej niż dwóch grup zaproponowane zostaną metody łączenia klasyfikatorów. Zaprezentowane metody klasyfikacji dla wielowymiarowych danych funkcjonalnych zostaną zilustrowane i omówione w kontekście analizy rzeczywistych, wielowymiarowych szeregów czasowych.

Literatura

[1] T. Górecki, M. Krzyśko, Ł. Waszak, *Functional Discriminant Coordinates,* Communications in Statistics - Theory and Methods 43, 1013–1025, 2014.

[2] L. Horváth, P. Kokoszka, *Inference for Functional Data with Applications,* Springer, New York, 2012.

[3] H.G. Müller, U. Stadtmüller, *Generalized functional linear models,* The Annals of Statistics, 33, 774–805, 2005.

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

Estymacja komponentów wariancyjnych w liniowych modelach mieszanych metodą największej wiarogodności: podejście algebraiczne

Mariusz Grządziel

Katedra Matematyki, Uniwersytet Przyrodniczy we Wrocławiu

Andrzej Michalski

Katedra Matematyki, Uniwersytet Przyrodniczy we Wrocławiu

W normalnych modelach liniowych mieszanych funkcja wiarogodności, a także funkcja resztowej wiarogodności (ang. *residual maximum likelihood* — w skrócie REML), może mieć więcej niż jedno maksimum lokalne, por. [1, rozdz. 7] oraz [2]. Zagadnienia obliczania ocen największej wiarogodności oraz ocen REML komponentów wariancyjnych w tych modelach należą zatem do klasy problemów optymalizacji niewypukłej. W referacie zostanie pokazane, że wyznaczanie ocen REML w przypadku, gdy liczba komponentów wariancyjnych jest równa dwa, może być sprowadzone do znajdowania wszystkich pierwiastków rzeczywistych odpowiednio zdefiniowanego wielomianu (podobnie, jak to ma miejsce w przypadku ocen największej wiarogodności). Omówiona zostanie również efektywność metod algebraicznych przedstawionych w rozdziale 6-tym pracy Gross i in. [1], które mogą być stosowane do obliczania ocen największej wiarogodności oraz ocen REML w modelach z więcej niż dwoma komponentami wariancyjnymi.

Literatura

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

[2] L. Henn, J. Hodges, *Multiple Local Maxima in Restricted Likelihoods and Posterior Distributions for Mixed Linear Models,* International Statistical Review 82, 90-105.

Charakteryzacje rozkładów Weibulla

Maria Iwińska, Magdalena Szymkowiak

Instytut Matematyki, Politechnika Poznańska

W pracy zostaną przedstawione charakteryzacje rozkładów Weibulla. Do charakteryzacji będzie wykorzystana intensywność starzenia. Ponadto zostanie zbadany porządek intensywności starzenia dla scharakteryzowanych rozkładów Weibulla.

Literatura

[1] Almalki, S.J., Nadarajah, S., *Modifications of the Weibull distribution: A review,* Reliability Engineering and System Safety, 124, 32–55, 2014.

[2] Bhattacharjee, S., Nanda, A.K., Misra, S.Kr., *Reliability analysis using ageing intensity function,* Statistics and Probability Letters, 83, 1364–1371, 2013.

[3] Nanda, A.K., Bhattacharjee, S., Alam, S.S., *Properties of aging intensity function,* Statistics and Probability Letters, 77, 365–373, 2007.

Prezentowane wyniki badań, zrealizowane w ramach tematu nr 04/43/DSPB/0082, zostały sfinansowane z dotacji na naukę przyznanej przez Ministerstwo Nauki i Szkolnictwa Wyższego.

Oszacowania wariancji statystyk porządkowych pochodzących z modeli urnowych

Krzysztof Jasiński

Wydział Matematyki i Informatyki, Uniwersytet Mikołaja Kopernika w Toruniu

Rozważamy modele losowania ze zwracaniem i bez zwracania. W obu przypadkach podajemy oszacowania wariancji statystyk porządkowych wyrażone w jednostkach wariancji pojedynczej obserwacji. Przedstawiamy także warunki ich osiągalności.

Literatura

[1] K. Jasiński, T. Rychlik, *Inequalities for variances of order statistics coming from urn models,* in review

Predykcja bayesowska dla podwójnie stochastycznego procesu Poissona

A. Jokiel-Rokita, D. Lazar, R. Magiera

Wydział Podstawowych Problemów Techniki, Politechnika Wrocławska

Rozpatruje się problem predykcji sumarycznej wielkości znaków w określonym przedziale czasowym w modelu znakowanego procesu punktowego, zadanym przez podwójnie stochastyczny proces Poissona N. Zakłada się, że N jest dwuwymiarowym procesem Poissona o losowej mierze intensywności $P \times \Theta$, gdzie P jest znaną miarą deterministyczną, a Θ – miarą losową. W rozważanym problemie zastosowano podejście bayesowskie zakładając, że Θ jest procesem gamma o znanych parametrach. Skonstruowano predyktor bayesowski przy zapobiegawczej funkcji straty. Zbadano wpływ parametru k, opisującego kształt założonej funkcji straty, na wartości predyktora.

Literatura

[1] D. Blackwell, J.B. MacQueen, *Ferguson distributions via Pólya urn schemes,* Ann. Statist., 1, 353-355, 1973.

[2] J. Grandell, *Doubly stochastic Poisson processes,* University of Stockholm, Institute of Actuarial Mathematics and Mathematical Statistics, 1975.

[3] A. Jokiel-Rokita, D. Lazar, R. Magiera, *Bayesian prediction in doubly stochastic Poisson process,* Metrika, DOI 10.1007/s00184-014-0484-x,
URL http://dx.doi.org/10.1007/s00184-014-0484-x.

[4] W. Niemiro, *Bayesian prediction with an asymmetric criterion in a nonparametric model of insurence risk,* Statistics, 40(4), 353-363, 2006.

[5] J.G. Norstrøm, *The use of precautionary loss function in risk analysis,* IEEE Transactions on Reliability, 45(3), 400-403, 1996.

Rozkłady chwil zatrzymania w pewnych sekwencyjnych procedurach estymacji

Alicja Jokiel-Rokita, Ryszard Magiera

Wydział Podstawowych Problemów Techniki, Politechnika Wrocławska

W wielu praktycznych problemach, np. w analizie niezawodności, czy w badaniach klinicznych, dane dostępne są jedynie w chwilach losowych. Załóżmy zatem, że obserwujemy jednocześnie n jednakowych obiektów pracujących niezależnie, które w losowych chwilach ulegają np. awarii. W tych losowych chwilach obserwujemy realizacje niezależnych zmiennych losowych $Y_1, \ldots, Y_n$ o jednakowym rozkładzie $P_\vartheta \in \mathcal{P}$, gdzie $\mathcal{P}$ jest daną parametryczną rodziną rozkładów. Załóżmy ponadto, że rozkład P_ϑ jest niezależny od rozkładu awarii. Na przykład w pracach Starra, Wardropa i Woodroofe'a (1976), Magiery (1982), Jokiel-Rokity i Magiery (2010) podano rozwiązanie problemów sekwencyjnej estymacji nieznanego parametru ϑ w pewnych modelach statystycznych na podstawie tak otrzymywanych danych, które nazywane są w literaturze danymi opóźnionymi. Pomimo że w pracach tych rozpatrywano różne rodziny rozkładów $\mathcal{P}$ i różne postaci funkcji straty związanej z błędem estymacji, to uzyskana w nich optymalna chwila zatrzymania τ_0 może być wyrażona pewną ogólną formułą.

W referacie przedstawimy rozkład i wartość oczekiwaną optymalnej chwili zatrzymania τ_0 oraz rozkład i wartość oczekiwaną liczby obserwacji w momencie zatrzymania (Jokiel-Rokita i Magiera (2014)). W oparciu o wyznaczone rozkłady, dokonamy porównania ryzyka optymalnej procedury sekwencyjnej i adaptacyjnej procedury sekwencyjnej zaproponowanej w pracy Starra, Wardropa i Woodroofe'a (1976).

Literatura

[1] A. Jokiel-Rokita, R. Magiera, *Estimation procedures with delayed observations,* JSPI, 140(4), 992-1002, 2010.

[2] A. Jokiel-Rokita, R. Magiera, *Distributions of stopping times in some sequential estimation procedures,* Metrika, 77, 617-634, 2014.

[3] R. Magiera, *Estimation with delayed observations,* Appl. Math., 17(2), 249-258, 1982.

[4] N. Starr, R. Wardrop, M. Woodroofe, *Estimating a mean from delayed observations,* Z. Wahrscheinlichkeitstheorie und Verw. Geb., 35:103–113, 1976.

Klasyfikacja i selekcja zmiennych w regresji logistycznej metodą losowych podprzestrzeni

Joanna Karłowska-Pik

Wydział Matematyki i Informatyki, Uniwersytet Mikołaja Kopernika w Toruniu

Klasyfikacja w przypadku wysoko wymiarowej przestrzeni zmiennych objaśniających i ograniczonej liczby obserwacji (tzw. *small n large p problem*) jest bardzo ważnym wyzwaniem w naukach biologicznych, a zwłaszcza w genomice i proteomice. Celem badań jest nie tylko konstrukcja efektywnych klasyfikatorów, ale przede wszystkim selekcja cech najbardziej informatywnych.
Dla zagadnienia regresji liniowej odpowiednie rozwiązanie zostało podane przez Mielniczuka i Teisseyre'a (2014) w oparciu o zmodyfikowaną metodę losowych podprzestrzeni (*RSM*). Polega ona na losowym wyborze pewnego zestawu zmiennych o zmniejszonej liczebności, rozwiązaniu wyjściowego problemu na tym zmniejszonym podzbiorze oraz przypisaniu zmiennym wag zależnych od ich ważności w utworzonym modelu. Procedura jest powtarzana, a wagi są agregowane. Otrzymane w ten sposób wartości wag stanowią podstawę do oceny ważności zmiennych w końcowym modelu.
Celem referatu jest przedstawienie rezultatów osiągniętych w wyniku zastosowania metody losowych podprzestrzeni dla regresji logistycznej, gdzie wagi są budowane w oparciu o wartości statystyki Walda. Szczególna uwaga zostanie zwrócona na specyficzne wymagania związane z użyciem tej statystyki oraz prezentację rozwiązań, które potrafią uczynić wspomnianą metodę bardziej efektywną. Regresja logistyczna z wykorzystaniem RSM jest konkurencyjną metodą dla algorytmu MCFS (Dramiński i inni (2008)).

Literatura

[1] Dramiński M., Rada-Iglesias A., Enroth S., Wadelius C., Koronacki J., Komorowski J., *Monte Carlo feature selection for supervised classification,* Bioinformatics, Vol. 24 no. 1 (2008), pp. 110–117.

[2] Mielniczuk J., Teisseyre P., *Using random subspace method for prediction and variable importance in linear regression,* Computational Statistics & Data Analysis, Vol. 71 (2014), pp. 725–742.

Miary rozproszenia dla danych wielowymiarowych

Adam Kołacz[1] **i Przemysław Grzegorzewski**[1,2]

[1]Wydział Matematyki i Nauk Informacyjnych, Politechnika Warszawska
[2]Instytut Badań Systemowych, Polska Akademia Nauk

W literaturze poświęconej statystyce opisowej można znaleźć wiele miar rozproszenia dla próbek złożonych z obserwacji jednowymiarowych (por. np. [1-3, 8]). W referacie zostanie przedstawiona ogólna definicja miar rozproszenia dla próbek utworzonych przez obserwacje k-wymiarowe ($k \geqslant 1$), tzn. $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^k$, a następnie zostanie podana pewna naturalna taksonomia tych miar.
W dalszej części zostanie zaproponowanych kilka miar rozproszenia dla próbek złożonych z obserwacji k-wymiarowych ($k \geqslant 2$). Do konstrukcji niektórych spośród tych miar wykorzystano multiodległości (por. [5-7]).

Literatura

[1] Bickel P.J., Lehmann E.L., *Descriptive statistics for nonparametric models (III. Dispersion),* Ann. Statist., 4, 1139-1158, 1976.

[2] Bickel P.J., Lehmann E.L., *Descriptive statistics for nonparametric models (IV. Spread),* w: J. Jureckova (Ed.), Contributions to Statistics (Academia, Prague), pp. 33-40, 1979.

[3] Gagolewski M., *Spread measures and their relation to aggregation functions,* European Journal of Operational Research, 2014 (w druku).

[4] Hyndman R.J., Fan, Y., *Sample quantiles in statistical packages,* American Statistician, 50, 361–365, 1996.

[5] Martín J., Mayor G., *How separated Palma, Inca and Manacor are?,* Proceedigns of the AGOP 2009, pp. 195-200, 2009.

[6] Martín J., Mayor G., *Multi-argument distances,* Fuzzy Sets and Systems, 167, 92-100, 2011.

[7] Martín J., Mayor G., Valero O., *Functionally expressible multidistances,* w: Proceedings of the EUSFLAT-LFA 2011, Atlantis Press, pp. 41-46, 2011.

[8] Wilcox R.R., *Introduction to Robust Estimation and Hypothesis Testing,* Elsevier, 2005.

Ograniczenia na wartości oczekiwane L-statystyk wyrażone przez średnią różnicę Gini'ego

Paweł Marcin Kozyra, Tomasz Rychlik

Instytut Matematyczny, Polska Akademia Nauk

Opisujemy tutaj metodę obliczania dolnych i górnych ograniczeń wartości oczekiwanych ustalonych, scentrowanych L-statystyk wyrażonych w jednostkach średniej różnicy Gini'ego pierwotnych niezależnych obserwacji o tym samym rozkładzie. Precyzyjne wartości ograniczeń są wyprowadzone dla pojedynczych statystyk pozycyjnych, ich różnic, jak również pewnych szczególnych przykładów L-statystyk. Prezentujemy także rodziny dyskretnych rozkładów, które osiągają te ograniczenia, możliwie w granicy.

Literatura

[1] Jones, M.C., Balakrishnan, N., *How are moments and moments of spacings related to distribution function,* J. Statist. Plann. Inference 103, 377–390, (2002)

[2] Rychlik, T., *Bounds for expectations of L-estimates. In: ,* Order Statistics: Theory & Methods (N. Balakrishnan, and C.R. Rao, eds.), Handbook of Statistics, vol. 16, North-Holland, Amsterdam, pp. 105–145, (1998)

[3] Rychlik, T., *Optimal bounds on L-statistics based on samples drawn with replacement from finite populations,* Statistics 38, 391–412, (2004)

[4] Rychlik, T., *Bounds on expectations of small order statistics from decreasing density populations,* Metrika 70, 369–381, (2009a)

[5] Rychlik, T., *Non-positive upper bounds on expectations of low rank order statistics from DFR populations,* Statistics 43, 53–63, (2009b)

[6] Rychlik, T., *Tight evaluations for expectations of small order statistics from symmetric and symmetric unimodal populations,* Statist. Probab. Lett. 79, 1488–1493, (2009c)

O wielomianach Charliera w kontekście weryfikacji poissonowskości rozkładu obserwacji

Teresa Ledwina i Grzegorz Wyłupek

Instytut Matematyczny, Polska Akademia Nauk
oraz Instytut Matematyczny, Uniwersytet Wrocławski

Rozważać będziemy klasyczny problem testowania zgodności rozkładu obserwacji z modelem Poissona

$$f(x;\lambda) = \frac{\lambda^x}{x!}e^{-\lambda}, \quad x \in N_0 = \{0,1,2,\ldots\}, \quad \lambda > 0.$$

W 1905 roku szwedzki astronom Charlier zaproponował następujący sposób modelowania rozkładów dyskretnych $f(x)$ na N_0

$$f(x) = f(x;\lambda)\left\{1 + \sum_{r=1}^{\infty} \frac{c_r}{r!} C_r(x;\lambda)\right\},$$

gdzie c_r są rzeczywistymi współczynnikami a

$$C_r(x;\lambda) = \frac{\partial^r f(x;\lambda)}{\partial \lambda^r} \cdot \frac{1}{f(x;\lambda)}.$$

Wielomiany $\{C_r\}_{r\geqslant 1}$ są ortogonalne i noszą nazwę wielomianów Charliera lub Poissona-Charliera. Ich ortonormalna wersja ma postać

$$h_r(x;\lambda) = \frac{1}{\sqrt{\lambda^r r!}} C_r(x;\lambda), \quad r \geqslant 1, \quad h_0(x;\lambda) \equiv 1, \quad x \in N_0.$$

Przeskalowane (przez $\sqrt{n}$) empiryczne współczynniki Fouriera w układzie $\{h_r\}_{r\geqslant 1}$, wyznaczone z próby $X_1,\ldots,X_n$, wyrażają się wzorem $U_r(\lambda) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} h_r(X_i;\lambda)$ a statystyki $U_r^2(\hat{\lambda})$, $r \geqslant 2$, $\hat{\lambda} = \bar{X}$ noszą nazwę gładkich komponent. Przy $r = 1$ zachodzi $U_1^2(\hat{\lambda}) = 0$.

W referacie przedyskutujemy dwa aspekty użycia tak zdefiniowanych komponent: selekcję alternatyw w badaniach mocy istniejących testów oraz konstrukcję nowych testów zgodności z rozkładem Poissona.

Dostateczność statystyki kwadratowo dostatecznej w modelu mieszanym

Augustyn Markiewicz

Katedra Metod Matematycznych i Statystycznych,
Uniwersytet Przyrodniczy w Poznaniu

Pojęcie dostateczności liniowej zostało wprowadzone przez Baksalarego i Kalę [1] i Drygasa [2] a pojęcie dostateczności kwadratowej przez Muellera [3], w obu wypadkach w kontekście stałego modelu liniowego. Przy założeniu normalności pokazano w [3], że statystyka kwadratowo dostateczna jest dostateczna. Celem pracy jest zbadanie własności tej statystyki gdy niektóre parametry okazują się być losowe, tzn. gdy poprawnym jest mieszany model liniowy. Pokazano, że przy założeniu normalności zachowuje ona własność dostateczności. Przedstawiono zastosowanie uzyskanego wyniku w zagadnieniu redukcji danych.

Literatura

[1] J. K. Baksalary, R. Kala, *Linear transformations preserving the best linear unbiased estimator in a general Gauss-Markoff model,* Annals of Statistics, 25, 423-463, 1981

[2] H. Drygas, *Sufficiency and completeness in the general Gauss-Markov model,* Sankhyā Ser A, 45, 88-98, 1983

[3] J. Mueller, *Sufficiency and completeness in the linear model,* Journal of Multivariate Analysis, 21, 312-323, 1987

Algorytm Metropolisa dla sieci bayesowskich z czasem ciągłym

Błażej Miasojedow

Instytut Matematyki Stosowanej i Mechaniki, Uniwersytet Warszawski

Sieci bayesowskie z czasem ciągłym (CTBN) są to procesy Markowa z czasem ciągłym na dyskretnej przestrzeni stanów, w których zależności pomiędzy składowymi mogą być przedstawione w postaci grafu Schwedera. Proces Markowa $X(t) = (X_1(t), \ldots, X_n(t))$ jest CTBN, jeśli intensywności przejść

$$Q(x, y) = \lim_{h \to 0} \frac{1}{h} \mathbb{P}\left(X(t+h) = y | X(t) = x\right) \text{ dla } x \neq y$$

można przedstawić w postaci

$$Q(x, y) = Q_i(x_i, y_i | x_{\mathrm{pa}(i)}) \text{ dla } x_i \neq y_i \text{ i } x_j = y_j \text{ dla wszystkich } j \neq i,$$

gdzie $x_{\mathrm{pa}(i)}$ oznacza konfigurację stanów w komponentach będących rodzicami x_i w grafie Schwedera. Ważnym z punktu widzenia wnioskowania statystycznego, a zarazem trudnym do symuacji rozkładem jest rozkład procesu $X(t)$ przy znanych częściowych obserwacjach, np. obserwowany jest proces tylko dla niektórych wierzchołków grafu lub obserwowany są stany procesu $X(t_1), ..., X(t_k)$ w ustalonych punktach czasowych $t_1, ..., t_k$.
W referacie przedstawiony będzie algorytm Metropolisa pozwalający na efektywne losowanie z powyższego rozkładu. Przedstawione będą wyniki symulacyjne oraz wyniki dotyczące tempa zbieżnosci zaproponowanych metod. Wyniki są oparte na wspólenj pracy z W. Niemiro, J. Noble oraz K. Opalskim.

Literatura

[1] BM, W. Niemiro, J. Noble oraz K. Opalski, *Metropolis-type algorithms for Continuous Time Bayesian Networks, preprint 2014.,*

Estymacja komponentów wariancyjnych w liniowych modelach mieszanych metodą największej wiarogodności: istnienie ocen parametrów

Andrzej Michalski

Katedra Matematyki, Uniwersytet Przyrodniczy we Wrocławiu

Mariusz Grządziel

Katedra Matematyki, Uniwersytet Przyrodniczy we Wrocławiu

W pracy rozważany jest problem istnienia ocen parametrów otrzymywanych metodą największej wiarogodności (ML) oraz największej resztowej wiarogodności (ang. *residual maximum likelihood*— w skr. REML) parametrów w mieszanym modelu liniowym. W literaturze statystycznej znane są prace (np. [2]), w których autorzy zajmujący się problemem istnienia ocen typu ML lub REML w modelach z komponentami wariancyjnymi powołują się na kluczowe twierdzenie z pracy Demidenki i Massam [1, Tw. 3.1] o istnieniu ocen ML parametrów modelu. Przeanalizowano dowód Demidenki i Massam tego twierdzenia, podającego warunek konieczny i dostateczny istnienia tych ocen. W dowodzie tego twierdzenia wskazano istotne błędy, a następnie je poprawiono. Ponadto przedstawiono nowy dowód Twierdzenia 3.4 sformułowanego przez Demidenkę i Massam [1], dotyczącego istnienia ocen typu REML.

Literatura

[1] E. Demidenko, H. Massam, *On the existence of the maximum likelihood estimate in variance components models,* Sankhyā: The Indian Journal of Statistics 61, 431–443, 1999.

[2] D. Birkes, S. S. Wulff, *Existence of the maximum likelihood estimates in normal variance-components models,* Journal of Statistics Planning and Inference 113, 35–47, 2003.

Oszacowania ilorazów rozproszeń mieszanek uporządkowanych rozkładów

Patryk Miziuła

Wydział Matematyki i Informatyki, Uniwersytet Mikołaja Kopernika w Toruniu
oraz Instytut Matematyczny, Polska Akademia Nauk

Rozważam model, w którym mieszankę tworzą nieznane rozkłady, wiadomy jest natomiast rozkład mieszający. Jest to sytuacja spotykana w wielu dziedzinach matematyki stosowanej, np. teorii niezawodności czy ubezpieczeniach. Interesują mnie miary rozproszenia takich mieszanek (np. wariancja, średnie bezwzględne odchylenie od mediany i wiele innych). W czasie prezentacji podam optymalne oszacowania ilorazów miar rozproszeń mieszanek, zależne jedynie od rozkładów mieszających. Wspomnę również o próbach optymalnego szacowania różnic miar położeń (np. wartości oczekiwanych, kwantyli) mieszanek tego typu.

Literatura

[1] P. Miziuła, T. Rychlik, *Extreme dispersions of semicoherent and mixed system lifetimes,* Journal of Applied Probability, przyjęta do druku.

[2] P. Miziuła, T. Rychlik, *Sharp bounds for lifetime variances of reliability systems with exchangeable components,* IEEE Transactions on Reliability, przyjęta do druku.

Kawałkami deterministyczne procesy Markowa i algorytmy MCMC

Wojciech Niemiro

Wydział Matematyki, Informatyki i Mechaniki, Uniwersytet Warszawski oraz
Wydział Matematyki i Informatyki, Uniwersytet Mikołaja Kopernika w Toruniu

Algorytm Metropolisa, pierwszy i najważniejszy z algorytmów MCMC, polega na generowaniu łańcucha Markowa odwracalnego względem docelowego rozkładu prawdopodobieństwa. Aby osiągnąć odwracalność, odrzuca się niektóre z proponowanych ruchów. Michel, Kapfer i Krauth (*op. cit.*) skonstruowali algorytm oparty na idei „maksymalnie nieodwracalnego" łańcucha Markowa, który nigdy nie odrzuca proponowanych ruchów. Zastosowanie tego algorytmu jest jednak ograniczone do pewnego modelu fizyki statystycznej, uwzględniającego oddziaływania parami. W naszej pracy pokażemy, jak skonstruować „maksymalnie nieodwracalny" algorytm dla dowolnego rozkładu docelowego na przestrzeni ciągłej. Wykorzystujemy w tym celu proces Markowa z czasem ciągłym, który ewoluuje deterministycznie pomiędzy losowymi momentami „zdarzeń". W momencie „zdarzenia" proces skokowo zmienia kierunek ruchu. Heurystyczne argumenty, związane z twierdzeniem Peskuna, pozwalają przypuszczać, że jest to droga do zwiększenia efektywności algorytmów MCMC. Wyniki symulacji, na razie przeprowadzonych w niewielkiej skali, są obiecujące. Wskazują na większą efektywność nowej metody w porównaniu z algorytmem Metropolisa.
Referat zawiera wyniki wspólnej pracy Błażeja Miasojedowa, Wojciecha Niemiro, Johna Noble i Krzysztofa Opalskiego.

Literatura

[1] Michel, M.; Kapfer, S.C.; Krauth, W., *Generalised Event-Chain Monte Carlo: Construction Rejection-Free Global-Balance Algorithms from Infinitesimal Steps.,* J. Chem. Phys. 140, 054116, 2014.

Ocena dokładności współczynników modelu opartego na pomiarach stanu nawierzchni dróg

Karol R. Opara

Instytut Badań Systemowych, Polska Akademia Nauk

Zabiegi związane z zimowym utrzymaniem dróg przyspieszają rozwój uszkodzeń ich nawierzchni. Zjawisku temu przeciwdziałać można poprzez stosowanie dodatków białkowo-ligninowych do soli drogowej, zmniejszających ilość ubytków kruszywa z nawierzchni asfaltowych. Ocenę skuteczności tej technologii oparto na badaniach terenowych w czterech europejskich państwach: Holandii, Szwecji, Danii i Austrii. Wyniki pomiarów przeanalizowano za pomocą modeli liniowych. Ocena dokładności wyznaczenia ich współczynników wymaga modyfikacji standardowych metod. Rezultaty pomiarów drogowych można bowiem agregować do odcinków diagnostycznych o długości od kilku do kilkuset metrów. Wybór ten wpływa na liczność próby statystycznej, zaburzając ocenę dokładności wyznaczenia współczynników modelu. Problem ten rozwiązano wprowadzając metodę blokowego bootstrapu. Umożliwia ona określenie dokładności współczynników jak również długości odcinków diagnostycznych, które można uznać za niezależne. Kolejną trudnością w opracowaniu oraz interpretacji modelu okazała się współliniowość zmiennych objaśniających. Jej skutki zredukowano poprzez regularyzację oraz ograniczenie dopuszczalnych wartości współczynników.

Literatura

[1] R. M. L. Demmer, W. Maslow i E. R. de Jong, *Deicing composition,* Wniosek patentowy nr PCT/EP2012/060542 zgłoszony przez Akzo Nobel, World Intellectual Property Organization, 2012.

[2] M. Skakuj i in., *Observation of the Development of Pavement Damages Over the Winter — Field Experiment,* Raport techniczny HELLER Ingenieurgesellschaft mbH, Darmstadt, Niemcy, 2014.

Modelowanie wieku pacjentów za pomocą częstości metylacji cytozyny w DNA - wyniki na podstawie danych z sekwencjonowania nowej generacji

Agnieszka Prochenka

Instytut Podstaw Informatyki, Polska Akademia Nauk

Metylacja to reakcja biochemiczna, która polega na przyłączaniu grup metylowych do cytozyny w DNA. Proces ten ma związek ze starzeniem się organizmu, a także występowaniem chorób takich jak nowotwory. Zagadnienie budowania modeli liniowych dla wieku opierając się na częstości metylacji jest szeroko opisane w literaturze. Autorom Hannum, Gregory, et al. w swojej pracy udało się zbudować model na podstawie metylacji zaledwie 71 cytozyn, dla którego pierwiastek średniego błędu kwadratowego na próbie testowej to około 5 lat. Pomimo rozwoju techniki sekwencjonowania nowej generacji, brak w literaturze prac, gdzie używa się jej w opisanym kontekście.
Podczas referatu przedstawię wyniki otrzymane we współpracy z Zakładem Genetyki Medycznej Warszawskiego Uniwersytetu Medycznego dotyczące modelowania wieku pacjentów za pomocą częstości metylacji cytozyny na podstawie największego w Polsce zbioru danych otrzymanych za pomocą techniki sekwencjonowania nowej generacji, obejmującego 214 pacjentów.

Literatura

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

Regresja rangowa z karą LASSO oraz nierówności z wyrocznią

Wojciech Rejchel

Wydział Matematyki i Informatyki, Uniwersytet Mikołaja Kopernika w Toruniu
oraz Wydział Matematyki, Informatyki i Mechaniki, Uniwersytet Wrocławski

Regresja rangowa związana jest z zadaniami, w których należy przewidzieć lub rozpoznać uporządkowanie obiektów na podstawie obserwacji zmiennych pomocniczych, opisujących cechy tych obiektów. Rozważamy estymatory regresji rangowej, które minimalizują ryzyko empiryczne z wypukłą funkcją straty i, często używaną w problemach wielowymiarowych, karą LASSO [1]. Ryzyko estymatorów szacujemy, używając "nierówności z wyrocznią". Wyniki te są ugólnieniem rezultatów dotyczących procesów empirycznych z karą LASSO [2,3] na U-procesy.

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] B. Tarigan, S. van de Geer, *Classifiers of support vector machine type with l_1 penalty,* Bernoulli 12, pp. 1045-1076, 2006.

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

Bootstrapowe predykcyjne obszary i przedziały ufności w modelach szeregów czasowych z zastosowaniem estymacji nieparametrycznej i semiparametrycznej.

Roman Różański

Wydział Podstawowych Problemów Techniki, Politechnika Wrocławska.

Zostanie przedstawione zastoswoanie metody bootstrapu sitowego (sieve bootsrap) do konstrukcji predykcyjnych przedziałów i obszarów ufności dla szerokiej klasy modeli szeregów czasowych. Bootstrap sitowy, wprowadzony i rozwijany przez P. Buehlmanna (P. Buehlmann (1998), (2002)), jest oparty w swej podstawowej idei na aproksymacji modelu szeregu czasowego ciągiem skończenie wymiarowych liniowych nodeli szeregów czasowych w nawiązaniu do nieparametrycznej estymacji metodą sita Grenandera (U. Grenander (1981)). Konstrukcja bootstrapowych, predykcyjnych obszarów i przedziałów ufności zostanie przeprowadzona dla modeli stacjonarnych szeregów czasowych z nieznanym trendem z zastosowaniem estymatorów nieparametrycznych i semiparametrycznych, wielowymiarowych, stacjonarnych szeregów czasowych oraz hetroskedastycznych modeli stacjonarnych szreegów czasowych z zakłóceniami typu GARCH.

Literatura

[1] P. Buehlmann, *Sieve bootstrap for smoothing in nonstationary time series,* The Annals of Statistics, vol 26, no 1, 48 - 83, 1998.

[2] P. Buehlmann, *Bootstraps for time series,* Statistical Science, vol. 17, no 1, 52 -72, 2002.

[3] G. Chłapiński and R. Różański, *Prediction intervals for time series models with trend via sieve bootstrap,* Journal of Statistical Planning and Inference, 143, 221-236, 2013.

[4] M. Kawecki and R. Różański, *Consistent semiparametric estimators in nonlinear egression models with dependent noise. Application to semiparametric trend estimation and bootstrap prediction intervals for time series,* preprint, 2014.

[5] R. Różański, G. Chłapiński,M. Hławka,K. Jamróz,M. Kawecki, A. Zagdański, *Prediction intervals and regions for multivariate time series modelsvia sieve bootstrap,* wysłane do publikacji, 2014.

[6] U. Grenander, *Abstract Inference,* Wiley, New York, 1981.

[7] R. Różański and A. Zagdański, *On the consistency of sieve bootstrap prediction intervals for stationary time series,* Discussiones Mathematicae. Probability and Statistics, Vol. 24, 5-40, 2004.

[8] A. Zagdański, *On the construction and properties of bootstrap-t prediction intervals for stationary time series,* Probability and Mathematical Statistics, 25, 133-153, 2005.

Optymalne oszacowania wartości oczekiwanych statystyk pozycyjnych z nieparametrycznych rodzin rozkładów generowanych przez porzadek wypukły

Tomasz Rychlik

Instytut Matematyczny, Polska Akademia Nauk

Załóżmy, że $X_1, \ldots, X_n$ są niezależnymi zmiennymi losowymi o jednakowej dystrybuancie F, która poprzedza ustaloną dystrybuantę W w porządku wypukłym. W szczególności, jeśli W jest dystrybuantą jednostajną lub wykładniczą, to F ma odpowiednio rosnącą gęstość i rosnącą intensywność awarii. Dla rodziny wszystkich rozkładów populacji poprzedzających daną dystrybuantę W w porządku wypukłym, przedstawimy precyzyjne górne oszacowania wartości oczekiwanych pojedynczych statystyk pozycyjnych i spacji (różnic sąsiednich statystyk pozycyjnych) opartych na $X_1, \ldots, X_n$ i wyrażonych w terminach średniej i odchylenia standardowego populacji. Scharakteryzujemy rozkłady, które osiągają równość w oszacowaniach oraz podamy szczegółowe wyniki dla klasy rozkładów o rosnącej gęstości.

Układy E-optymalne przy skorelowanych błędach losowych

Łukasz Smaga

Wydział Matematyki i Informatyki,
Uniwersytet im. Adama Mickiewicza w Poznaniu

W referacie rozważane są E-optymalne oraz E-efektywne układy o macierzach, których elementy należą do zbioru $\{-1, 0, 1\}$. O błędach losowych zakładamy, że mają takie same wariancje oraz równe dodatnie kowariancje. Takie układy odpowiadają chemicznym układom wagowym oraz doświadczeniom czynnikowym typu 3^p. Udowodniona została E-optymalność układów spełniających pewne warunki, gdy liczba obserwacji $n \equiv 2 \pmod 4$. Zaprezentowano również konstrukcje układów E-optymalnych wykorzystujące macierze $\mathbf{S}_n$ oraz macierze Hadamarda. Skonstruowane układy E-optymalne nie muszą być ani A- ani D-optymalne. Ponadto, w przypadkach, gdy układy E-optymalne nie są znane, podane zostały układy o wysokiej E-efektywności.

Literatura

[1] M. Jacroux, C.S. Wong, J.C. Masaro, *On the optimality of chemical balance weighing designs,* Journal of Statistical Planning and Inference, 8, 231-240, 1983.

[2] A.W. Marshall, I. Olkin, B.C. Arnold, *Inequalities: Theory of Majorization and Its Applications,* Second Edition, Springer Series in Statistics, 2011.

[3] J. Masaro, C.S. Wong, *Robustness of optimal designs for correlated random variables,* Linear Algebra and its Applications, 429, 1639-1646, 2008a.

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

[5] J. Masaro, C.S. Wong, *Robustness of A-optimal designs,* Linear Algebra and its Applications, 429, 1392-1408, 2008c.

[6] Ł. Smaga, *Uniquely E-optimal designs with* $n \equiv 2$ *(mod* 4*) correlated observations,* Linear Algebra and its Applications (przyjęta do druku)
http://dx.doi.org/10.1016/j.laa.2014.08.022

Subspace clustering – nowa metoda iteracyjna

Piotr Sobczyk

Wydział Podstawowych Problemów Techniki, Politechnika Wrocławska

W pracy zaproponowano nową metodę klasteryzacji wielowymiarowych danych przy założeniu, że zmienne są zgrupowane w nisko wymiarowe podprzestrzenie. Jest ona rozszerzeniem algorytmu k-średnich, gdzie rolę środków przyjmują składowe główne, a miara podobieństwa oparta jest o korelację. Jakość uzykiwanych klastrów została porównana w symulacjach z istniejącymi metodami. Dodatkowo stworzono procedurę estymacji liczby klastrów i ich wymiarów w opariu o kryterium BIC. Metoda została zaimplementowana w pakiecie w języku R.

Literatura

[1] M. Chavent, B. Liquet, V. Kuentz-Simonet, J. Saracco, *ClustOfVar: An R Package for the Clustering of Variables,* Journal of Statistical Software, 2012.

[2] R. Vidal, *Subspace clustering,* IEEE Signal Processing Magazine, 2011

[3] M. Verbanck, J. Josse, and F. Husson, *Regularised pca to denoise and visualise data,* Statistics and Computing, pages 1–16, 2013.

[4] P. Sobczyk, *varclust: Variables clustering,* R package, 2014

Algorytmy typu boosting w modelowaniu różnicowym

Michał Sołtys

Instytut Podstaw Informatyki, Polska Akademia Nauk

Modelowanie różnicowe (*uplift modeling*) to poddziedzina uczenia maszynowego, która zajmuje się przewidywaniem wpływu określonego działania na jednostkę. Działaniem tym może być procedura medyczna, akcja marketingowa lub inny czynnik kontrolowany przez człowieka. Model różnicowy wymaga dwóch zbiorów testowych: grupy badanej, poddanej działaniu czynnika, i kontrolnej, wolnej od jego wpływu. Celem modelowania jest oszacowanie zmiany prawdopodobieństwa pożądanej reakcji jednostki pod wpływem czynnika.

Algorytmy typu boosting są dobrze znanymi metodami wzmacniania modeli klasyfikacyjnych. Ich adaptacja do problemu różnicowego może pozwolić na znaczną poprawę modelu bazowego.

W referacie przedstawione zostaną konstrukcje różnych wariantów boostingu różnicowego z omówieniem porównań eksperymentalnych proponowanych metod i algorytmów bazowych.

Literatura

[1] M. Sołtys, S. Jaroszewicz, P. Rzepakowski, *Ensemble methods for uplift modeling,* Data Mining and Knowledge Discovery, 1–29, 2014.

[2] N.J. Radcliffe and P.D. Surry, *Real-world uplift modelling with significance-based uplift trees,* Portrait Technical Report TR-2011-1, Stochastic Solutions, 2011.

[3] Y. Freund, R.E. Schapire, *A decision-theoretic generalization of on-line learning and an application to boosting,* Journal of Computer and System Sciences, 55(1):119–139, 1997.

O agregacji indywidualnych decyzji w wielokryterialnych zadaniach optymalnego stopowania

Krzysztof Szajowski

Instytut Matematyki i Informatyki, Politechnika Wrocławska

Przedmiotem rozważań są modele wielokryterialnego optymalnego zatrzymania wprowadzone przez Kurano, Nakagami i Yasudę (1980) dla ciągów niezależnych wektorów losowych, uogólnone między innymi przez Fergusona (2005) oraz Szajowskiego i Yasudę (1996). Celem jest uwzględnienie niejednorodoności modelu zarówno w kryteriach jak i samym obserwowanym procesie. Uzyskane wyniki pozwalają wprowadzić dyskontowanie wypłat oraz koszt za obserwacje. Agregacja decyzji modelowana jest grą prostą.

Literatura

[1] Ferguson, T.S., *Selection by committee.,* In: Nowak, A., Szajowski, K. (eds.) Advances in dynamic games, Ann. Internat. Soc. Dynam. Games, vol. 7, pp. 203–209. Birkhäuser Boston, MA, 2005.

[2] Kurano, M., Yasuda, M., Nakagami, J., *Multi-variate stopping problem with a majority rule.,* J. Oper. Res. Soc. Jap. 23, 205–223, 1980.

[3] Szajowski, K., Yasuda, M., *Voting procedure on stopping games of Markov chain.,* In: A.H. Christer, Shunji Osaki and Lyn C. Thomas (eds.) UK-Japanese Res. Workshop on Stoch. Modelling in Innovative Manufecuring, July 21-22, 1995. LN in Econ. and Math. Sys., vol. 445, pp. 68–80. Moller Centre, Churchill College, Univ. Cambridge, UK, Springer, 1996.

O estymacji metodą minimalnego kontrastu w problemach odwrotnych

Zbigniew Szkutnik

Wydział Matematyki Stosowanej, Akademia Górniczo-Hutnicza w Krakowie

Omówione zostaną własności nieparametrycznych estymatorów funkcyjnego parametru $f \in \mathcal{F} \subset L_2(\mathbb{R}^d)$ w liniowym problemie odwrotnym, gdy obserwowane są w pewnej przestrzeni $\mathcal{Y}$ niezależne kopie elementu losowego o gęstości Af względem μ, z pewnym operatorem $A : L_2(\mathbb{R}^d) \to L_2(\mathcal{Y}, \mu)$. Minimaksowe tempa zbieżności zależą od wielkości klasy $\mathcal{F}$ i od własności operatora, a ich osiągnięcie wymaga zwykle zastosowania przy konstrukcji estymatora jakiejś formy regularyzacji. W przypadku estymatorów minimalnego kontrastu regularyzacja może polegać np. na minimalizacji empirycznego kontrastu na δ-sieciach, zamiast na całej klasie $\mathcal{F}$. Wiadomo, że bez regularyzacji estymatory minimalnego kontrastu mają zwykle suboptymalne tempa zbieżności, gdy klasa $\mathcal{F}$ jest zbyt bogata, co wyraża się np. zbyt szybkim wzrostem funkcji entropii metrycznej w sąsiedztwie zera (Birgé, Massart (1993)). Dlatego zaskakuje niedawny wynik Klemelä i Mammena (2010) głoszący osiąganie minimaksowych temp zbieżności przez estymatory minimalizujące empiryczną funkcję kontrastu na całej klasie $\mathcal{F}$ bez założeń ograniczających entropię. Okazuje się, że ich dowód ma istotną lukę, której wypełnienie wymaga jednak nałożenia odpowiednich warunków na entropię metryczną klasy $\mathcal{F}$.

Literatura

[1] L. Birgé, P. Massart, *Rates of convergence for minimum contrast estimators,* Probab. Theory Relat. Fields, 97, 113-150, 1993.

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

Zgodność zmodyfikowanej wersji Bayesowkiego kryterium informacyjnego przy nienormalnym błędzie

Piotr Szulc

Instytut Matematyki i Informatyki, Politechnika Wrocławska

W referacie omówię problem wyboru istotnych zmiennych w modelu rzadkiej regresji przy pomocy zmodyfikowanej wersji Bayesowkiego kryterium informacyjnego. Przedstawię twierdzenie o zgodności tego kryterium bez założenia, że błąd losowy pochodzi z rozkładu normalnego.

Zgodność selekcji Uogólnionego Kryterium Informacyjnego dla wysokowymiarowej regresji logistycznej.

Hubert Szymanowski

Instytut Podstaw Informatyki, Polska Akademia Nauk

Prezentujemy moetdę selekcji cech w regresji logistycznej polegającą na wyborze podzbioru zmiennych, który minimalizuje Uogólnione Kryterium Informacyjne (GIC), spośród wszystkich podzbiorów zmiennych o rozmiarze nieprzekraczającym k. Dowodzimy zgodności tej reguły przy ogólnych założeniach, rozszerzając w ten sposób wyniki z pracy Chen Chen (2012). Rozważamy znacznie szerszą rodzinę funkcji kryterialnych oraz przypadek gdy k zależy od liczby obserwacji. Wyniki obejmują sytuację gdy liczba predyktorów rośnie wykładniczo względem liczby obserwacji.

Literatura

[1] Bogdan M, Doerge R and Ghosh J, *Modifying the Schwarz Bayesian information criterion to locate multiple interacting quantitative trait loci.,* Genetics, 167:989-999, 2004.

[2] Chen J and Chen Z, *Extended Bayesian Information Criteria for model selection with large model spaces.,* Biometrika, 95:759-771, 2008.

[3] Chen J and Chen Z , *Extended BIC for small-n-large-p sparse GLM.,* Statist Sinica, 22:555-574, 2012.

[4] Luo S and Chen Z, *Extended BIC for linear regression models with diverging number of relevant features and high or ultra-high feature spaces.,* J Statist Plann Inference, 143:494-504, 2013.

[5] Pokarowski P and Mielniczuk J, *Combined ℓ_1 and greedy ℓ_0 least squares for linear model selection.,* submitted, 2014.

Charakteryzacje odwróconych rozkładów Weibulla

Magdalena Szymkowiak, Maria Iwińska

Instytut Matematyki, Politechnika Poznańska

W pracy zostaną przedstawione niektóre własności tzw. odwrotnej intensywności starzenia. Będzie ona wykorzystana do charakteryzacji odwróconych rozkładów Weibulla. Ponadto zostanie zdefiniowany i zbadany porządek odwrotnej intensywności starzenia dla scharakteryzowanych odwróconych rozkładów Weibulla.

Literatura

[1] Almalki, S.J., Nadarajah, S., *Modifications of the Weibull distribution: A review,* Reliability Engineering and System Safety, 124, 32–55, 2014.

[2] Bhattacharjee, S., Nanda, A.K., Misra, S.Kr., *Reliability analysis using ageing intensity function,* Statistics and Probability Letters, 83, 1364–1371, 2013.

[3] Finkelstein, M.S., *On the reversed hazard rate,* Reliability Engineering and System Safety, 78, 71–75, 2002.

Prezentowane wyniki badań, zrealizowane w ramach tematu nr 04/43/DSPB/0079, zostały sfinansowane z dotacji na naukę przyznanej przez Ministerstwo Nauki i Szkolnictwa Wyższego.

Wybrane metody klasyfikacji wieloetykietowej

Paweł Teisseyre

Instytut Podstaw Informatyki, Polska Akademia Nauk

Klasyfikacja wieloetykietowa (ang. multi-label classification) odnosi się do sytuacji w której każdy z obserwowanych obiektów jest opisany wektorem zmiennych objaśniających oraz wektorem K odpowiedzi (najczęściej binarnych) zwanych etykietami. Na podstawie danych uczących buduje się model który umożliwia przewidywanie etykiet dla nowych obiektów opisanych pewnymi cechami. W ostatnich kilku latach problem klasyfikacji wieloetykietowej wzbudził bardzo duże zainteresowanie. Metody klasyfikacji wieloetykietowej znajdują zastosowanie w takich dziedzinach jak: kategoryzacja tekstów, automatyczna anotacja obrazów, genomika i wiele innych. W referacie przedstawię popularne rozwiązania: metodę BR (ang. binary relevance), LP (ang. label powerset), łańcuchy klasyfikatorów (ang. classifier chains) oraz model Isinga. Przedstawię pewne własności teoretyczne modelu opartego na łańcuchu klasyfikatorów w którym pojedyńczy klasyfikator jest modelem logistycznym. W referacie przedstawię również wyniki eskperymentów symulacyjnych przeprowadzonych na rzeczywistych zbiorach danych.

Literatura

[1] G. Tsoumakas, I. Katakis, *Multi-label classification: An overview,* International Journal of Data Warehousing and Mining, 3, 1-13, 2007.

[2] Dembczyński, et. al., *On label dependence and loss minimization in multi-label classification,* Machine Learning, 88, 5-45, 2012.

[3] Read, et. al, *Classifier chains for multi-label classification,* Proceedings of the 20th European Conference on Machine Learning, 254-269, 2009

[4] Madjarov, et. al, *An extensive experimental comparison of methods for multi-label learning,* Pattern Recognition, 45, 3084-3104, 2012

[5] Cheng, et. al., *Sparse Ising Models with Covariates,* manuscript, 2012

Semiparametryczna estymacja krzywej ROC

Rafał Topolnicki, Alicja Jokiel-Rokita

Wydział Podstawowych Problemów Techniki, Politechnika Wrocławska

Wystąpienie będzie dotyczyć estymacji krzywej ROC w podejściu semiparametrycznym na przykładzie modelu binormalnego [1]. Rozważone zostaną trzy metody estymacji semiparametrycznej spotykane w literaturze – uogólniona metoda najmniejszych kwadratów, metoda najmniejszej odległości oraz pewna jej modyfikacja [2]. Zaproponowane i omówione zostaną również modyfikacje analizowanych metod semiparametrycznych polegające na zastąpieniu przy ich konstrukcji empirycznej krzywej ROC przez ciągłe nieparametryczne estymatory. Otrzymane w modelu binormalnym estymatory zostaną porównane na drodze symulacji numerycznych z szeregiem estymatorów nieparametrycznych. Ostatnia część referatu będzie poświęcona analizie rzeczywistych danych.

Literatura

[1] F. Hsieh, B. W. Turnbull, *Nonparametric and semiparametric estimation of the receiver operating curve,* Ann. Statist., 24, 25-40, 1996.

[2] O. Davidov, Y. Nov, *Improving an estimator of Hsieh and Turnbull for the binormal ROC curve,* J. Statist. Plann. Inference, 142, 872-877, 2012.

[3] A. Jokiel-Rokita, M. Pulit, *Nonparametric estimation of the ROC curve based on smoothed empirical distribution functions,* Stat. Comput., 23, 703-712, 2013.

Testowanie parasolkowatych alternatyw

Grzegorz Wyłupek

Instytut Matematyczny, Uniwersytet Wrocławski

Nieparametryczny problem testowania parasolkowatych alternatyw, w literaturze, zwykle formułowany jest w następujący sposób

$$\mathcal{H}_0 : F_1 = \cdots = F_p = \cdots = F_k,$$
$$\mathcal{A}_u : \text{istnienie trendu } F_1 \geqslant \cdots \geqslant F_p \leqslant \cdots \leqslant F_k,\ F_i \neq F_j$$
$$\text{dla pewnych } 1 \leqslant i < j \leqslant k.$$

Punkt p nazywany jest szczytem parasolki. Rozważane są dwa przypadki, gdy p jest znane lub też nie. Alternatywy tego typu są najczęściej rozpatrywane w rozmaitych badaniach efektywności.
Takie postawienie problemu nie uwzględnia jednak sytuacji, gdy porządek z $\mathcal{A}_u$ nie zachodzi, a $\mathcal{H}_0$ nie jest prawdziwa. Problem ten jest istotny, a szczególną uwagę zwrócono na niego w pracy Terpstry i Magel (2003) formułując przy tym postulat o konieczności kontroli funkcji mocy na całym zbiorze rozkładów z hipotezy. Problem został dostrzeżony również przez praktyków (por. Nakas i Alonzo, 2007 oraz Alonzo i inni, 2009).
Wobec powyższego, planuje się bardziej bezpieczne postawienie problemu. Mianowicie, będziemy weryfikować

$$\mathcal{H} : \text{brak trendu} \quad F_1 \geqslant \ldots \geqslant F_p \leqslant \ldots \leqslant F_k,\ F_i \neq F_j$$
$$\text{dla pewnych} \quad 1 \leqslant i < j \leqslant k,$$
$$\mathcal{A}_u : \text{istnienie trendu} \quad F_1 \geqslant \ldots \geqslant F_p \leqslant \ldots \leqslant F_k,\ F_i \neq F_j$$
$$\text{dla pewnych} \quad 1 \leqslant i < j \leqslant k.$$

W komunikacie przedstawimy konstrukcję nowej statystyki testowej w powyższym problemie. Przedyskutujemy uzyskane wyniki teoretyczne oraz zaprezentujemy wyniki przeprowadzonych badań symulacyjnych.

Literatura

[1] T. A. Alonzo, Ch. T. Nakas, C. T. Yiannoutsos, S. Bucher, *A comparison of tests for restricted orderings in the three-class case,* Statistics in Medicine, 28, 1144-1158, 2009.

[2] Ch. T. Nakas, T. A. Alonzo, *ROC graphs for assessing the ability of a diagnostic marker to detect three disease classes with an umbrella ordering,* Biometrics, 63, 603-609, 2007

[3] J. T. Terpstra, R. C. Magel, *A new nonparametric test for the ordered alternative problem,* Journal of Nonparametric Statistics, 15, 289-301, 2003.

Przegląd problemów i metod związanych z analiza danych dotyczących ekspresji genów na podstawie eksperymentu głębokiego sekwencjonowania

Joanna Zyprych-Walczak, Alicja Szabelska, Idzi Siatkowski

Katedra Metod Matematycznych i Statystycznych,
Uniwersytet Przyrodniczy w Poznaniu

W chwili obecnej dostępny jest szereg metod, które pozwalają na analizę ekspresji genów. Jedną z najnowocześniejszych technologii w tym kierunku jest sekwencjonowanie nowej generacji (RNA-seq). W poniższej prezentacji zostanie przedstawiony przebieg eksperymentu RNA-seq z głównym naciskiem położonym na statystyczną analizę danych. Zostaną zasygnalizowane istniejące problemy, z którymi eksperymentator może spotkać się na różnych etapach analiz. Zaprezentowane zostaną istniejące metody dotyczące ekspresji genów. W dalszej części przedstawione zostaną przykładowe analizy porównawcze, które będą miały na celu zaprezentowanie możliwe ścieżki podejmowania decyzji przez badacza.

Literatura

[1] S. Anders, W. Huber, *Differential expression analysis for sequence count data,* Genome Biol, 11(10):R106, 2010.

[2] M. Krzyśko, W. Wołyński, T. Górecki, M. Skorzybut, *Systemy uczące się. Rozpoznawanie wzorców analiza skupień i redukcja wymiarowości,* Wydawnictwo Naukowo Techniczne, 2008.

[3] A. Leśniewska, M. Okoniewski, *rnaSeqMap: a Bioconductor package for RNA sequencing data exploration,* BMC bioinformatics, 12:200, 2011.

[4] J. Li, R. Tibshirani, *Finding consistent patterns: a nonparametric approach for identifying differential expression in rna-seq data,* Statistical Methods in Medical Research, 2011.

[5] J. Li, D. Witten, I. Johnstone, R. Tibshirani, *Normalization, testing, and false discovery rate estimation for rna-sequencing data,* Biostatistics, 13(3):523–538, 2012.

[6] M. Okoniewski, A. Leśniewska, A. Szabelska, J. Zyprych-Walczak, M. Ryan, M. Wachtel, T. Morzy, B. Schäffer, R. Schlapbach, *Preferred analysis methods for single genomic regions in RNA sequencing revealed by processing the shape of coverage,* Nucleic acids research, December 2011.

[7] M. Robinson, D. McCarthy, G. Smyth, *edger: a bioconductor package for differential expression analysis of digital gene expression data,* Bioinformatics, 26(1):139—140, 2010.

[8] M. Robinson, A. Oshlack i in., *A scaling normalization method for differential expression analysis of rna-seq data,* Genome Biol, 11(3):R25, 2010.

Część IV

Lista uczestników

1. **prof. Tadeusz Bednarski**
   Wydział Prawa Administracji i Ekonomii, Uniwersytet Wrocławski
   t.bednarski@prawo.uni.wroc.pl
2. **dr Mariusz Bieniek**
   Instytut Matematyki, Uniwersytet Marii Curie-Skłodowskiej w Lublinie
   mariusz.bieniek@umcs.lublin.pl
3. **mgr Krzysztof Boczkowski**
   Wydział Podstawowych Problemów Techniki, Politechnika Wrocławska
   Krzysztof.Boczkowski@pwr.edu.pl
4. **dr hab. Małgorzata Bogdan**
   Wydział Podstawowych Problemów Techniki, Politechnika Wrocławska
   malgorzata.bogdan@pwr.edu.pl
5. **dr Katarzyna Brzozowska-Rup**
   Wydział Zarządzania i Modelowania Komputerowego, Politechnika Świętokrzyska
   brzozows@poczta.fm
6. **prof. Tadeusz Caliński**
   Katedra Metod Matematycznych i Statystycznych,
   Uniwersytet Przyrodniczy w Poznaniu
   calinski@up.poznan.pl
7. **dr Bogdan Ćmiel**
   Wydział Matematyki Stosowanej, Akademia Górniczo-Hutnicza w Krakowie
   cmielbog@gmail.com
8. **dr hab. Antoni Leon Dawidowicz**
   Wydział Matematyki i Informatyki, Uniwersytet Jagielloński
   antoni.leon.dawidowicz@im.uj.edu.pl
9. **dr Anna Dudek**
   Wydział Matematyki Stosowanej, Akademia Górniczo-Hutnicza w Krakowie
   aedudek@agh.edu.pl
10. **mgr Kamil Dyba**
    Wydział Matematyki i Informatyki, Uniwersytet Wrocławski
    kamil.dyba@math.uni.wroc.pl
11. **dr hab. Karol Dziedziul**
    Wydział Fizyki Technicznej i Matematyki Stosowanej, Politechnika Gdańska
    kdz@mif.pg.gda.pl
12. **dr Katarzyna Filipiak**
    Katedra Metod Matematycznych i Statystycznych,
    Uniwersytet Przyrodniczy w Poznaniu
    kasfil@up.poznan.pl
13. **dr Agnieszka Goroncy**
    Wydział Matematyki i Informatyki, Uniwersytet Mikołaja Kopernika w Toruniu
    gemini@mat.umk.pl
14. **dr Tomasz Górecki**
    Wydział Matematyki i Informatyki,
    Uniwersytet im. Adama Mickiewicza w Poznaniu
    tomasz.gorecki@amu.edu.pl

15. **dr Mariusz Grządziel**
Katedra Matematyki, Uniwersytet Przyrodniczy we Wrocławiu
mariusz.grzadziel@up.wroc.pl

16. **dr hab. Przemysław Grzegorzewski**
Wydział Matematyki i Nauk Informacyjnych, Politechnika Warszawska
pgrzeg@mini.pw.edu.pl

17. **dr Maria Iwińska**
Instytut Matematyki, Politechnika Poznańska
maria.iwinska@put.poznan.pl

18. **dr Krzysztof Jasiński**
Wydział Matematyki i Informatyki, Uniwersytet Mikołaja Kopernika w Toruniu
krzys@mat.umk.pl

19. **dr hab. Alicja Jokiel-Rokita**
Wydział Podstawowych Problemów Techniki, Politechnika Wrocławska
alicja.jokiel-rokita@pwr.edu.pl

20. **dr Joanna Karłowska-Pik**
Wydział Matematyki i Informatyki, Uniwersytet Mikołaja Kopernika w Toruniu
joanka@mat.umk.pl

21. **mgr Adam Kołacz**
Wydział Matematyki i Nauk Informacyjnych, Politechnika Warszawska
a.kolacz@mini.pw.edu.pl

22. **mgr Paweł Marcin Kozyra**
Instytut Matematyczny, Polska Akademia Nauk
pawel_m_kozyra@wp.pl

23. **prof. dr hab. Mirosław Krzyśko**
Wydział Matematyki i Informatyki,
Uniwersytet im. Adama Mickiewicza w Poznaniu
mkrzysko@amu.edu.pl

24. **dr Agnieszka Kulawik**
Wydział Matematyki, Fizyki i Chemii, Uniwersytet Śląski w Katowicach
agnieszka.kulawik@us.edu.pl

25. **mgr Daniel Lazar**
Wydział Podstawowych Problemów Techniki, Politechnika Wrocławska
daniel.lazar@pwr.edu.pl

26. **prof. Teresa Ledwina**
Instytut Matematyczny, Polska Akademia Nauk
ledwina@impan.pan.wroc.pl

27. **prof Ryszard Magiera**
Wydział Podstawowych Problemów Techniki, Politechnika Wrocławska
ryszard.magiera@pwr.edu.pl

28. **dr Piotr Majerski**
Wydział Matematyki Stosowanej, Akademia Górniczo-Hutnicza w Krakowie
majerski@agh.edu.pl

29. **prof. dr hab. Augustyn Markiewicz**
Katedra Metod Matematycznych i Statystycznych,
Uniwersytet Przyrodniczy w Poznaniu
amark@up.poznan.pl

30. **dr hab. Marek Męczarski**
Instytut Ekonometrii, Szkoła Główna Handlowa
mecz@sgh.waw.pl

31. **dr Błażej Miasojedow**
Instytut Matematyki Stosowanej i Mechaniki, Uniwersytet Warszawski
bmia@mimuw.edu.pl

32. **dr hab. Andrzej Michalski**
Katedra Matematyki, Uniwersytet Przyrodniczy we Wrocławiu
apm.mich@gmail.com

33. **mgr Patryk Miziuła**
Wydział Matematyki i Informatyki, Uniwersytet Mikołaja Kopernika w Toruniu
bua@mat.umk.pl

34. **prof. Wojciech Niemiro**
Wydział Matematyki, Informatyki i Mechaniki, Uniwersytet Warszawski oraz Wydział Matematyki i Informatyki, Uniwersytet Mikołaja Kopernika w Toruniu
wniemiro@gmail.com

35. **prof. Magdalena Niewiadomska-Bugaj**
Department of Statistics, Western Michigan University
m.bugaj@wmich.edu

36. **dr Konrad Nosek**
Wydział Matematyki Stosowanej, Akademia Górniczo-Hutnicza w Krakowie
konosek@agh.edu.pl

37. **mgr Karol Opara**
Instytut Badań Systemowych, Polska Akademia Nauk
karol.opara@ibspan.waw.pl

38. **mgr Agnieszka Prochenka**
Instytut Podstaw Informatyki, Polska Akademia Nauk
a.prochenka@phd.ipipan.waw.pl

39. **dr Wojciech Rejchel**
Wydział Matematyki i Informatyki, Uniwersytet Mikołaja Kopernika w Toruniu
iggypop@mat.umk.pl

40. **dr hab. Roman Różański**
Wydział Podstawowych Problemów Techniki, Politechnika Wrocławska
Roman.Rozanski@pwr.edu.pl

41. **prof. Tomasz Rychlik**
Instytut Matematyczny, Polska Akademia Nauk
trychlik@impan.pl

42. **dr Łukasz Smaga**
Wydział Matematyki i Informatyki,
Uniwersytet im. Adama Mickiewicza w Poznaniu
ls@amu.edu.pl

43. **mgr Piotr Sobczyk**
Wydział Podstawowych Problemów Techniki, Politechnika Wrocławska
Piotr.Sobczyk@pwr.edu.pl

44. **mgr Michał Sołtys**
Instytut Podstaw Informatyki, Polska Akademia Nauk
m.soltys@phd.ipipan.waw.pl

45. **dr Alicja Szabelska**
Katedra Metod Matematycznych i Statystycznych,
Uniwersytet Przyrodniczy w Poznaniu
aszab@up.poznan.pl

46. **prof. Krzysztof Szajowski**
Instytut Matematyki i Informatyki, Politechnika Wrocławska
krzysztof.szajowski@pwr.edu.pl

47. **dr hab. Zbigniew Szkutnik**
Wydział Matematyki Stosowanej, Akademia Górniczo-Hutnicza w Krakowie
szkutnik@agh.edu.pl

48. **mgr Piotr Szulc**
Instytut Matematyki i Informatyki, Politechnika Wrocławska
piotr.a.szulc@pwr.edu.pl

49. **mgr Hubert Szymanowski**
Instytut Podstaw Informatyki, Polska Akademia Nauk
h.szymanowski@phd.ipipan.waw.pl

50. **dr Magdalena Szymkowiak**
Instytut Matematyki, Politechnika Poznańska
magdalena.szymkowiak@put.poznan.pl

51. **dr Paweł Teisseyre**
Instytut Podstaw Informatyki, Polska Akademia Nauk
teisseyrep@ipipan.waw.pl

52. **mgr Rafał Topolnicki**
Wydział Podstawowych Problemów Techniki, Politechnika Wrocławska
rafal.topolnicki@pwr.edu.pl

53. **dr Irena Wistuba**
Wydział Matematyki, Fizyki i Chemii, Uniwersytet Śląski w Katowicach
iwistuba@poczta.onet.pl

54. **mgr Jakub Wojdyła**
Wydział Matematyki Stosowanej, Akademia Górniczo-Hutnicza w Krakowie
jwojdyla@agh.edu.pl

55. **dr Grzegorz Wyłupek**
Instytut Matematyczny, Uniwersytet Wrocławski
wylupek@math.uni.wroc.pl

56. **dr Joanna Zyprych-Walczak**
Katedra Metod Matematycznych i Statystycznych,
Uniwersytet Przyrodniczy w Poznaniu
zjoanna@up.poznan.pl

Spis treści

1. Część I
Przedmowa **3**

2. Część II
Program konferencji **7**

3. Część III
Streszczenia **15**
Wykład wiodący **17**
Referaty **21**

4. Część IV
Lista uczestników **65**