

XLVI Konferencja

Statystyka Matematyczna

Lublin

30 listopada – 2 grudnia 2020

Organizatorzy

Komisja Statystyki Matematycznej Komitetu Matematyki PAN
Instytut Matematyki UMCS w Lublinie

Sponsorzy

Instytut Matematyki UMCS

Komitet Naukowy

prof. Małgorzata Bogdan, UWr
prof. Jan Mielniczuk, IPI PAN/MINI PW

Komitet Organizacyjny

dr hab. Mariusz Bieniek, prof. UMCS
dr hab. Monika Budzyńska, prof. UMCS
dr Małgorzata Cudna, IM UMCS
dr Aleksander Kowalski, IM UMCS

Redaktor

Mariusz Bieniek

Spis treści

1	Program konferencji	4
2	Zaproszeni wykładowcy	5
3	Streszczenia	8

Program konferencji

Zaproszeni wykładowcy

Nieparametryczna estymacja prawdopodobieństwa ruiny

Lesław Gajek

Politechnika Łódzka

Techniczna ruina (niewypłacalność) zakładu ubezpieczeń jest zjawiskiem rzadko obserwowanym w rzeczywistości, dlatego do estymacji prawdopodobieństwa ruiny stosuje się wyrafinowane wnioskowania oparte na znajomości modeli probabilistycznych działalności zakładów ubezpieczeń. Modele przełącznikowe uogólniają wiele znanych modeli ryzyka w ubezpieczeniach (zobacz na przykład Lu (2006)). Wykorzystany jest w nich jednorodny łańcuch Markowa, który skacząc pomiędzy stanami łańcucha, zmienia rozkład wielkości szkody lub/i czasu oczekiwania na nią. Z kolei ubezpieczyciel może dostosować swoją strategię do zmienionego stanu procesu Markowa zmieniając wysokość składki ubezpieczeniowej. W pracy Gajek, Rudź (2018) pokazano, że wektor prawdopodobieństw ruiny Ψ w nieskończonym horyzoncie czasu jest punktem stałym wektorowego operatora ryzyka w takim modelu. Pokażemy konsekwencje tego faktu dla rzędu zbieżności estymatorów prawdopodobieństwa ruiny. Omówimy słabe i mocne strony estymatorów podanych w literaturze oraz zaproponujemy nową postać estymatora. W tym celu skorzystamy z własności, że operator ryzyka jest kontrakcją w odpowiedniej przestrzeni metrycznej. Podamy konstrukcję empirycznego odpowiednika operatora ryzyka, a następnie zdefiniujemy jego punkt stały Ψ_n jako estymator wektora Ψ . Pokażemy, że estymator Ψ_n jest zbieżny do Ψ w odpowiedniej metryce. Wyniki zilustrujemy na przykładach numerycznych.

Literatura

- [1] Y. Lu (2006), *On the severity of ruin in a Markov-modulated risk model*, Scandinavian Actuarial Journal, 4, 183-202
- [2] L. Gajek, M. Rudź (2018), *Banach Contraction Principle and ruin probabilities in regime switching-models*, Insurance: Mathematics and Economics, 80, 45-53

Wprowadzenie do metody bootstrap

Anna Dudek

Akademia Górniczo-Hutnicza

Wykład poświęcony będzie metodzie bootstrapu zaproponowanej przez Bradleya Efrona w 1979. Jest to metoda resamplingowa, służąca do aproksymacji nieznanego rozkładu statystyki. Dowodem ogromnego sukcesu tej metody, jest przyznanie za nią Bradleyowi Efronowi w 2019 roku międzynarodowej nagrody w dziedzinie statystyki *International Prize in Statistics*.

Podczas prezentacji zostanie przedstawiona idea metody bootstrap oraz jej wersje dla danych niezależnych, stacjonarnych i pewnych klas niestacjonarnych szeregów czasowych. Omówione zostaną metody konstrukcji bootstrapowych przedziałów ufności. Ponadto przedstawione zostaną najważniejsze wyniki dotyczące zgodności bootstrapu. Zobrazowana zostanie uniwersalność tego podejścia i szerokie spektrum możliwych zastosowań. Jednocześnie podane zostaną przykłady braku zgodności bootstrapu.

Literatura

- [1] A. C. Davison, D. V. Hinkley (1997), *Bootstrap Methods and their Application*, Cambridge University Press
- [2] B. Efron, R.J. Tibshirani (1993), *An Introduction to the Bootstrap*, Chapman and Hall, New York
- [3] P. Hall (1992), *The Bootstrap and Edgeworth Expansion*, Springer
- [4] S.N. Lahiri (2003), *Resampling Methods for Dependent Data*, Springer, New York

Streszczenia

Wybór optymalnej kombinacji liniowej dwóch statystyk porządkowych przy estymacji kwantyli

Luiza Pańczyk, Mariusz Bieniek

Instytut Matematyki UMCS w Lublinie

Niech $X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$ oznaczają statystyki porządkowe z próby rozmiaru n . Rozważamy problem estymacji kwantyli rozkładu populacji $F^{-1}(p)$ ustalonego rzędu $p \in (0, 1)$ przy użyciu kombinacji liniowych dwóch sąsiednich statystyk porządkowych postaci $(1 - \alpha)X_{j:n} + \alpha X_{j+1:n}$. Dla zadanego rozmiaru próby i rzędu kwantyla szukamy takich wartości j oraz α , dla których estymator kwantyla powyższej postaci jest optymalny w klasie wszystkich rozkładów ze skończoną wariancją. Intuicyjnie jasne jest, że j powinno być bliskie np , więc tradycyjnie najczęściej jako j wybiera się $[np]$ lub $[np] + 1$. Ponadto, zwykle α jest wybierane jako liniowa funkcja np .

Zastosujemy nowe kryterium optymalności oparte na osiągalnych oszacowaniach obciążenia $EX_{j:n} - F^{-1}(p)$. Pokażemy, że optymalny wybór dla α nie pokrywa się z tradycyjnym wyborem oraz porównamy uzyskany estymator z klasycznymi estymatorami kwantyli (zob. Hyndman, Fan (1996)).

Literatura

[1] R. J. Hyndman, Y. Fan (1996), *Sample Quantiles in Statistical Packages*, The American Statistician 50, 361-365

Odporna estymacja bayesowska przy funkcji straty Bregmana i rodzinach rozkładów a priori opartych na funkcjach zniekształcających

Agata Boratyńska

Kolegium Analiz Ekonomicznych
Szkola Główna Handlowa SGH

Rozważamy problem estymacji parametru rzeczywistego w bayesowskim modelu statystycznym przy niepełnej informacji a priori. Niedokładność w specyfikacji rozkładu a priori modelowana jest za pomocą rodziny rozkładów a priori, które (w porządku opartym na ilorazie wiarygodności) leżą między dwoma rozkładami opartymi na wybranych funkcjach zniekształcających zadany rozkład a priori. Przy uogólnionej funkcji straty Bregmana (obejmującej szereg znanych funkcji strat, m.in. kwadratową, ważoną kwadratową funkcję straty, LINEX, uogólnioną funkcję straty Steina) wyznaczamy oscylację estymatora bayesowskiego i estymatory o Γ -minimaksowej utracie a posteriori.

Literatura

- [1] J.P. Arias-Nicolás, F. Ruggeri, A. Suárez-Llorens (2016), *New Classes of Priors Based on Stochastic Orders and Distortion Functions*, Bayesian Analysis 11, 1107–1136
- [2] A. Karimnezhad and A. Parsian (2018), *Bayesian and Robust Bayesian Analysis in a General Setting*, Commun in Statist. Theory and Methods 48, 3899-3920

Network Analysis of Commuting Flows in Poland. A Comparative Static Approach

Mariola Chrzanowska and Konrad Furmańczyk

Warsaw University of Life Sciences

Commuting patterns have a fundamental role in understanding socio-economic systems. It is a habitual act of leaving one's home and travelling to work. It is an important spatial equilibrating mechanism in the labour market [1]. According to the closed-economy market model, commuting reduces disparities in regional labour market outcomes (e.g. unemployment rate and wage). Over recent decades an increase in the number of commuters (i.e. people living in one municipality and working in others) has been observed. A moderate increase in the distance covered by these employees can be noticed as well. In recent years, more workers have been travelling from suburbs to city centres and between suburbs [3]. Concentration of economic activity in fewer locations is one of the reasons for commuting [4]. The main goal of this paper is to identify and analyse relations between poviats from the perspective of the labour market, especially commuting over the years. The research includes comparing the following aspects: topology of networks, structures of commuters and intensity of commuting. We compare graphs' topology for several years (using power law degree and basic characteristics of networks). The structure of commuters is described by a weighted directed graph. In consequence, we examine the difference between the structures of these graphs over the years using a permutation test. The intensity of commuting is explained by logistic regression. Information in the commute matrix format was obtained from the POLTAX database, where data from the tax system is stored. The data came from PIT-11 / 8B and PIT-40 forms submitted by employers to tax offices in Poland (2005, 2010, 2015).

Literatura

[1] D. Persyn and W. Torfs (2015), *A gravity equation for commuting with an application to estimating regional border effects in Belgium.*, Journal

of Economic Geography, 16(1), 155–175

[2] G. J. Borjas (2001), *Does Immigration Grease the Wheels of the Labor Market?* *Brookings Papers on Economic Activity*, Economic Studies Program, The Brookings Institution, vol. 32(1), 69-134

[3] M. Pohanka and S. Fitzgerald (2004), *Urban sprawl and you: how sprawl adversely affects worker health*, *AAOHN J*;52(6):242-246

[4] P. Krugman (1991), *Increasing Returns and Economic Geography*, *Journal of Political Economy* 99 (3), 483-499

Estymacja metodą największej wiarygodności na podstawie próby cenzurowanej z rozkładu dyskretnego

Anna Dembińska

Wydział Matematyki i Nauk
Informacyjnych, Politechnika Warszawska

Referat poświęcony będzie parametrycznej estymacji metodą największej wiarygodności na podstawie dyskretnych danych cenzurowanych. Rozważymy cenzurowanie typu drugiego, tzn. takie, w którym do eksperymentu bierzemy n elementów o czasach życia będących zmiennymi losowymi iid i rejestrujemy czasy awarii tych elementów, ale jedynie do chwili r -tej ($r < n$) awarii. Podamy ogólne warunki na parametryczną rodzinę rozkładów czasu życia pojedynczego elementu, które gwarantują, że z prawdopodobieństwem 1 istnieje mocno zgodny ciąg estymatorów największej wiarygodności. Następnie pokażemy, że w przypadku trzech rozkładów używanych do modelowania dyskretnych czasów życia - rozkładu Poissona, dwumianowego i ujemnego dwumianowego – estymatory największej wiarygodności są wyznaczone jednoznacznie, jeśli tylko istnieją i wtedy są mocno zgodne.

Prezentowane wyniki zostały uzyskane wspólnie z Krzysztofem Jasińskim i są przedstawione w [1].

Literatura

[1] A. Dembińska, K. Jasiński (2020), *Maximum likelihood estimators based on discrete component lifetimes of a k -out-of- n system.*, TEST doi.org/10.1007/s11749-020-00724-0

O związku między rzutem a estymatorem największej wiarogodności macierzy kowariancji

Katarzyna Filipiak

Instytut Matematyki, Politechnika Poznańska

W modelach wielowymiarowych z liniową strukturą macierzy kowariancji i znaną wartością oczekiwaną Szatrowski (1980) pokazał, że estymatory największej wiarogodności parametrów macierzy kowariancji mają postać jawną wtedy i tylko wtedy gdy przestrzeń macierzy kowariancji jest podprzestrzenią kwadratową. Celem tej prezentacji jest rozszerzenie wyników Szatrowskiego na model z nieznaną wartością oczekiwaną oraz podanie warunków dostatecznych, aby rzut estymatora największej wiarogodności macierzy kowariancji nie posiadającej żadnej struktury na przestrzeń macierzy określonych dodatkowo o zadanej strukturze był estymatorem największej wiarogodności macierzy kowariancji o tej strukturze.

Literatura

- [1] K. Filipiak, M. John, A. Markiewicz (2020), *Comments on maximum likelihood estimation and projections under multivariate statistical models*, In: Holgersson, T., Singull, M. (Eds.) *Recent Developments in Multivariate and Random Matrix Analysis*, pp. 51-66. Springer
- [2] T.H. Szatrowski (1980), *Necessary and sufficient conditions for explicit solutions in the multivariate normal estimation problem for patterned means and covariances*, *Annals of Statistics* 8, 802–810

Testowanie własności MTP2 dla rozkładu normalnego

Konrad Furmańczyk

Instytut Informatyki Technicznej, Szkoła
Główna Gospodarstwa Wiejskiego w Warszawie

W referacie zostanie przedstawiony nowy test badania własności MTP2 [1] dla wektorów losowych o rozkładzie normalnym. Proponowany test jest oparty na procedurze multitestowania Bonferroniego [2]. Test ten wykorzystuje asymptotyczny rozkład normalny empirycznego estymatora macierzy precyzji w przypadku nisko-wymiarowym oraz korekty estymatora glosso [3] w przypadku wysoko-wymiarowym. Test ten zostanie porównany symulacyjnie z uniwersalnym testem [4], który może być również zastosowany do badania własności MTP2.

Literatura

- [1] S. Karlin and Y. Rinott (1980), *Classes of Orderings of Measures and Related Correlation Inequalities. Multivariate Totally Positive Distributions*, Journ Multiv Anal. 10, 467-498
- [2] K. Furmańczyk (2020), *A multitest procedure for testing MTP2 for Gaussian distribution*, preprint
- [3] S. van de Geer, S., P. Bühlmann, Y. Ritov, R. Dezeure (2014), *On asymptotically optimal confidence regions and tests for high-dimensional models*, Ann. Stat. 42(3):1166–1202
- [4] L. Wasserman, A. Ramdas, S. Balakrishnan (2020), *Universal Inference*, <https://arxiv.org/abs/1912.11436>

O górnych niedodatnich oszacowaniach funkcjonalów statystycznych z rozkładów o malejącej uogólnionej intensywności awarii

Agnieszka Goroncy

Wydział Matematyki i Informatyki
Uniwersytet Mikołaja Kopernika w Toruniu

Motywacją do badań była chęć rozwiązania problemu oszacowania z dołu wartości oczekiwanych standaryzowanych k -tych rekordów, $k \geq 1$, pochodzących z rodziny rozkładów o malejącej uogólnionej intensywności awarii ($DGFR(\alpha)$), która definiowana jest poprzez porządek wypukły dystrybuanty rozkładu wyjściowego względem dystrybuanty uogólnionego rozkładu Pareto (G_α , $\alpha \in \mathbb{R}$). W przypadku tych wartości rekordowych, dla których oszacowania dolne są niedodatnie, można je łatwo otrzymać korzystając z wyników Bieńka i Goroncy (2020). Problem analogicznych oszacowań dolnych nieujemnych można również rozwiązać, gdy rozważane są pewne szczególne klasy rozkładów z tej rodziny, dla $\alpha = 0$ (Goroncy, 2020) lub $\alpha = 1$ (Goroncy, 2017). Powstaje pytanie, czy analogicznie można rozważać nieujemne oszacowania dolne dla pozostałych $\alpha \neq \{0, 1\}$. Spróbujemy na nie odpowiedzieć, przedstawiając pewne uogólnienie wyników dotyczących górnych niedodatnich oszacowań dla średnich uogólnionych statystyk pozycyjnych pochodzących z pewnych szczególnych rodzin rozkładów: o malejącej gęstości (Goroncy, 2017) oraz malejącej intensywności awarii (Goroncy, 2020).

Literatura

- [1] Bieńka, M., Goroncy, A. (2020), *Sharp lower bounds on expectations of gOS based on $DGFR$ distributions*, Stat. Papers 61:3, 1027–1042
- [2] Goroncy, A. (2020), *On the upper bounds on expectations of gOS s based on DFR and $DFRA$ distributions*, Statistics, 54:2, 402–414
- [3] Goroncy, A. (2017), *Upper non-positive bounds on expectations of generalized order statistics from DD and DDA populations*, Comm. Statist.-

Theory Meth., 46:24, 11972–11987

BLOCK MATRIX APPROXIMATION VIA ENTROPY LOSS FUNCTION

Malwina Janiszewska, Augustyn
Markiewicz, Monika Mokrzycka

Poznan University of Life Sciences, Poznań

The purpose of this talk is to present a procedure of approximation of symmetric positive definite matrix by symmetric block partitioned matrices with structured off-diagonal blocks. We consider off-diagonal blocks which can get one of the following structures: all of its elements are equal (this block can be also considered as a part of compound symmetry structure) or the off-diagonal blocks are proportional to respective sub-block of the first-order autoregression matrix. The entropy loss function is chosen as approximation criterion. This procedure is applied in statistical problem of covariance structure identification. We perform simulation studies to verify whether the entropy loss function recognizes given covariance structure properly. The presented results were carried out in cooperation with A. Markiewicz and M. Mokrzycka.

Literatura

- [1] X. Cui, C. Lin, J. Zhao, L. Zeng, D. Zhang, J. Pan (2016), *Covariance structure regularization via Frobenius norm discrepancy*, Linear Algebra Appl. 510, 124–145
- [2] M. Janiszewska, A. Markiewicz, M Mokrzycka (2020), *Block matrix approximation via entropy loss function*, Applications of Mathematics 65(6), 829-844
- [3] L. Lin, N. J. Higham, J. Pan (2014), *Covariance structure regularization via entropy loss function*, Comput. Statist. Data Anal. 72, 315–327

O liczbie popsutych elementów w działającym systemie koherentnym

Krzysztof Jasiński

Wydział Matematyki i Informatyki,
Uniwersytet Mikołaja Kopernika w Toruniu

Rozważymy systemy k -spośród- n , dla których czasy życia komponentów mają rozkłady dyskretne. Policzymy prawdopodobieństwo, że jest dokładnie i , $i = 1, 2, \dots, n - k$, popsutych elementów pod warunkiem, że system działa w chwili t . Następnie rozszerzymy uzyskane wyniki na dowolny system koherentny.

Literatura

[1] K. Jasiński (2020), *The number of failed components in a coherent working system when the lifetimes are discretely distributed*, w recenzji

Estymacja bayesowska parametrów potęgowego procesu Weibulla

Alicja Jokiel-Rokita, Ryszard Magiera

Politechnika Wroclawska, Wydział Matematyki

Potęgowy proces Weibulla jest procesem odnowy z trendem, w którym funkcja trendu jest potęgowa, natomiast funkcja odnowy jest dystrybucją rozkładu Weibulla. W referacie przedstawimy propozycje nieinformujących rozkładów a priori trójwymiarowego parametru tego procesu. Omówimy wnioski z symulacji, których celem było porównanie dokładności estymatorów bayesowskich względem ustalonego nieinformującego rozkładu a priori przy kwadratowej funkcji straty i przy błędzie absolutnym z dokładnością estymatorów największej wiarygodności. Przedstawimy analizę danych rzeczywistych i propozycje sprawdzania poprawności przyjętego modelu.

Dokładne prawa wielkich liczb dla rozkładów typu asymetrycznego Pareto

Paweł Kurasiński

Instytut Matematyki, UMCS

Prezentacja będzie dotyczyć dokładnych praw wielkich liczb. Twierdzenia tego typu polegają na badaniu zbieżności prawie pewnej (ewentualnie innych typów zbieżności - według prawdopodobieństwa lub kompletnej) ważonych sum niezależnych zmiennych losowych o nieskończonej wartości oczekiwanej lub równej zero do pewnej niezerowej i skończonej stałej. W prezentacji pokażemy dpwl dla rozkładów typu asymetrycznego Pareto ze zbieżnością prawie pewną, według prawdopodobieństwa oraz kompletną. W literaturze były badane mocne oraz słabe dpwl dla rozkładu dwuogonowego Pareto oraz dla rozkładu asymetrycznego Cauchy'ego. W dalszej części pokażemy stosowność naszych twierdzeń do badania ważonych sum ilorazów niezależnych zmiennych losowych, które nie są konieczne dodatnie. We wcześniejszych pracach problem ten był badany dla jedynie dodatnich zmiennych losowych.

Literatura

- [1] P. Matuła, P. Kurasiński, A. Adler, H. Naderi (2020), *A Note on Exact Laws of Large Numbers for Asymmetric Pareto-type Distributions with Applications to Ratios of Random Variables*, Bulletin of the Institute of Mathematics Sinica New Series, Vol. 15, No, 2, pp. 177–186
- [2] A. Adler (2008), *Laws of large numbers for two tailed Pareto random variables*, Probab. Math. Statist. 28, 121–128
- [3] A. Adler (2007), *Laws of large numbers for asymmetrical Cauchy random variables*, J. Appl. Math. Stoch. Anal. Art. ID 56924, 6pp.

Estymatory ściąające dla funkcji masy prawdopodobieństwa

Małgorzata Łazęcka, Jan Mielniczuk

Wydział Matematyki i Nauk Informatycznych PW
Instytut Podstaw Informatyki PAN

Rozważane będą sposoby estymacji funkcji masy prawdopodobieństwa p (w przypadku pary zmiennych dyskretnych (X, Y) ozn. $p = p(x, y)$). Przedstawiony zostanie estymator ściąający typu Jamesa-Steina postaci

$$\hat{p}_\lambda = \lambda \hat{p}^{(1)} + (1 - \lambda) \hat{p}^{(2)},$$

gdzie $p^{(1)}$ jest estymatorem związanym z modelem ograniczonym i jego wariancja jest niższa niż wariancja $p^{(2)}$ (np. jest etymatorem największej wiarogodności w modelu, w którym zakładamy niezależność zmiennych X i Y lub rozkład jednostajny na (X, Y)), $p^{(2)}$ jest estymatorem nieobciążonym, zaś parametr $\lambda \in [0, 1]$ jest dobierany w taki sposób, by uzyskany estymator p_λ minimalizował błąd średniokwadratowy.

W trakcie referatu zostanie przedstawiona propozycja nowego estymatora p_λ , w którym parametr λ_{SE} minimalizuje błąd kwadratowy:

$$\begin{aligned} \lambda_{SE} &= \arg \min_{\lambda \in [0,1]} SE(\lambda) = \arg \min_{\lambda \in [0,1]} \sum_{x,y} (p(x, y) - \hat{p}_\lambda(x, y))^2 \\ &= \frac{\sum_{x,y} (\hat{p}^{(2)}(x, y) - p(x, y)) (\hat{p}^{(2)}(x, y) - \hat{p}^{(1)}(x, y))}{\sum_{x,y} (\hat{p}^{(1)}(x, y) - \hat{p}^{(2)}(x, y))^2}, \end{aligned}$$

pokazany zostanie sposób estymacji λ_{SE} oraz zastosowanie estymatora $\hat{p}_{\lambda_{SE}}$ do selekcji zmiennych opartej na miarach pochodzących z teorii informacji.

Literatura

[1] J. Hausser, K. Strimmer (2008), *Entropy Inference and the James-Stein Estimator, With Application to Nonlinear Gene Association Networks*, The Journal of Machine Learning Research 10, 1469–1484

[2] K. Sechidis, L. Azzimonti, A. Pocock, G. Corani, J. Weatherall, G. Brown (2019), *Efficient feature selection using shrinkage estimators*, Machine Learning 108, 1261–1286

Linear admissible estimation revisited

Augustyn Markiewicz

Poznań University of Life Sciences

The problem of admissible linear estimation under the Gauss-Markov model was extensively studied by Rao (1976). Rao's work stimulated further research in this area. Baksalary et al. (1992, 1995) studied this problem in a possibly singular model, while Baksalary and Mathew (1988) and Markiewicz (1998) studied it in a possibly misspecified model. Another subject of research considered by Markiewicz (1996) and Markiewicz and Puntanen (2009) was a specified subclass of admissible estimators that are in addition linearly sufficient. The purpose of this paper is to present the development of the theory and its state-of-the-art, along with some new supplementary results.

Literatura

- [1] J.K. Baksalary and T. Mathew (1988), *Admissible linear estimation in a general Gauss-Markov model with an incorrectly specified dispersion matrix*, Journal of Multivariate Analysis 27, 53–67
- [2] J.K. Baksalary, C.R. Rao, and A. Markiewicz (1992), *A study of the natural restrictions on the estimation problems in the singular Gauss-Markov model*, Journal of Statistical Planning and Inference 31, 335–351
- [3] J.K. Baksalary, A. Markiewicz, and C.R. Rao (1995), *Admissible linear estimation in the general Gauss-Markov model with respect to an arbitrary quadratic risk function*, Journal of Statistical Planning and Inference 44, 341–347
- [4] A. Markiewicz (1996), *Characterization of general ridge estimators*, Statistics & Probability Letters 27, 145–148
- [5] A. Markiewicz (1998), *Comparison of linear restricted models with respect to the validity of admissible and linearly sufficient estimators*, Statistics & Probability Letters 38, 347–354
- [6] A. Markiewicz and S. Puntanen (2009), *Admissibility and linear suf-*

efficiency in linear model with nuisance parameters, Statistical Papers, 50, 847—854

[7] C.R. Rao (1976), *Estimation of parameters in a linear model*, Annals of Statistics, 4, 1023–1037

Dokładne prawa wielkich liczb dla ilorazów zmiennych losowych i ich zastosowania

Przemysław Matuła

Instytut Matematyki, Uniwersytet Marii Curie–Skołodowskiej

W referacie omówimy ważne mocne prawa wielkich liczb dla zmiennych losowych o jednakowych i różnych rozkładach i o nieskończonej wartości oczekiwanej. Twierdzenia takie nazywane są w literaturze dokładnymi prawami wielkich liczb. Przedstawimy zastosowania podanych twierdzeń do badania asymptotyki ilorazów niezależnych zmiennych losowych oraz pewnych statystyk porządkowych.

Literatura

- [1] P. Matuła, P. Kurasiński and A. Adler (2019), *Exact strong laws of large numbers for ratios of the smallest order statistics*, Statist. Probab. Lett. 152, 69–73
- [2] P. Matuła, A. Adler and P. Kurasiński (2020), *On exact strong laws of large numbers for ratios of random variables and their applications*, Comm. Statist. Theory Methods 49, 3153–3167

Estymacja macierzy kowariancji o strukturze wstęgowej macierzy Toeplitza metodą shrinkage

Adam Mieldzioc

Uniwersytet Przyrodniczy w Poznaniu

Estymacja macierzy kowariancji odgrywa kluczową rolę w statystyce matematycznej. Znalezienie dobrze uwarunkowanego estymatora macierzy kowariancji o zadanej strukturze i dobrych własnościach statystycznych może być zadaniem trudnym i czasochłonnym obliczeniowo, zwłaszcza w przypadku dużej liczby obserwowanych cech. Dlatego w celu znalezienia estymatora o pożądanym własnościach i zadanej strukturze często wykorzystuje się metodę „shrinkage” z odpowiednią dobraną funkcją straty.

Estymatory macierzy kowariancji uzyskane w pracy zostaną porównane między innymi z estymatorami największej wiarygodności z ograniczeniem na współczynnik uwarunkowania macierzy (Won, Lim, Kim, Rajaratnam, 2013) w modelu, w którym badane jest wiele cech lub jedna cecha w wielu punktach czasowych.

Literatura

- [1] Y. Chen, A. Wiesel, Y.C. Eldar, A.O. Hero (2010), *Shrinkage Algorithms for MMSE Covariance Estimation*, IEEE Transactions on Signal Processing, 58:5016–5029
- [2] M. John and A. Mieldzioc (2019), *The comparison of the estimators of banded Toeplitz covariance structure under the high-dimensional multivariate model*, Communications in Statistics - Simulation and Computation
- [3] O. Ledoit and M. Wolf (2004), *A well-conditioned estimator for large-dimensional covariance matrices*, Journal of Multivariate Analysis 88(2), 365–411
- [4] JH. Won, J. Lim, SJ. Kim, B. Rajaratnam (2013), *Condition-Number-Regularized Covariance Estimation*, Journal of the Royal Statistical Society. Series B (Statistical Methodology) 75(3), 427–450

Estymacja w modelu logistycznym z częściową obserwowalnością typu PU

Jan Mielniczuk

Instytut Podstaw Informatyki PAN i Wydział
Matematyki i Nauk Informatycznych PW

W referacie rozważymy problem estymacji w modelu logistycznym dla pary zmiennych $(X, Y) \in R^p \times \{0, 1\}$, gdzie zmienna binarna Y jest nieobserwowalna i jest zastąpiona przez obserwowalną zmienną binarną S , przy czym $P(S = 1|Y = 1, X) = c < 1$ oraz $P(S = 0|Y = 0, X) = 1$ (model PU; *Positive Unlabelled*). Przedyskutowane zostanie metoda naiwna, polegająca na dopasowaniu modelu logistycznego do obserwacji z rozkładu $P_{X,S}$ oraz metoda największej wiarygodności dla jednoczesnej estymacji parametrów β modelu logistycznego i częstości etykietowania c . W szczególności pokazane zostaną związki metody naiwnej ze złą specyfikacją i własność ściągania parametru β w kierunku 0. Przedstawione zostaną również wyniki badań symulacyjnych dla estymacji prawdopodobieństwa aposteriori i c .

Wyniki uzyskano wspólnie z M. Łazęcką i P. Teisseyrem.

Literatura

- [1] J. Baker, J. Davis (2020), *Learning from positive and unlabelled data: a survey*, Machine Learning, 109, 719-760
- [2] P. Teisseyre, J. Mielniczuk, M.Łazęcka (2020), *Different strategies for fitting logistic regression for positive and unlabelled data*, Proceedings of the International Conference on Computational Science ICCS 2020, LNCS, 3-17

Aproksymacja macierzy kowariancji wybranymi strukturami w modelach podwójnie wielowymiarowych

Monika Mokrzycka

Instytut Genetyki Roślin PAN

Celem tego referatu jest przedstawienie najlepszego przybliżenia symetrycznej, określonej dodatnio macierzy Ω stopnia n przez określony dodatnio iloczyn Kroneckera macierzy Ψ i Σ , gdzie Ψ i Σ są dowolnymi macierzami symetrycznymi lub jeden ze składników iloczynu Kroneckera ma korelacyjną strukturę kompletnej symetrii lub autoregresji stopnia pierwszego. Jako miarę rozbieżności przyjęte zostały entropijna funkcja straty oraz kwadratowa funkcja straty. Przedstawione wyniki mogą być szeroko stosowane w statystyce wielowymiarowej, na przykład do identyfikacji struktury kowariancyjnej wielowymiarowych danych, do określenia estymatorów struktury kowariancji czy też do określenia odległości między hipotezą zerową i alternatywną w zaganieniu testowania hipotez o strukturach kowariancyjnych.

Prezentowane wyniki dotyczące entropijnej funkcji straty można znaleźć w pracach Filipiak i in. [1] oraz Filipiak i in. [2].

Literatura

- [1] K. Filipiak, D. Klein, M. Mokrzycka (2018), *Estimators comparison of separable covariance structure with one component as compound symmetry matrix*, Electronic Journal of Linear Algebra 33, 83–98
- [2] K. Filipiak, D. Klein, A. Markiewicz, M. Mokrzycka (2020), *Approximation with a Kronecker product structure with one component as compound symmetry or autoregression via entropy loss function*, Linear Algebra and its Applications, doi.org/10.1016/j.laa.2020.10.013

Planowanie eksperymentów symulacyjnych do oceny metaheurystyk

Karol Opara

Instytut Badań Systemowych PAN

Metaheurystyki, takie jak algorytmy ewolucyjne czy rojowe, stanowią jedne z najskuteczniejszych metod optymalizacji globalnej w dziedzinie ciągłej. Różnorodność i nieregularność klas funkcji niewypukłych czy wielomodalnych skutkuje brakiem punktu odniesienia w postaci algorytmów dokładnych. Do oceny dokładności oraz szybkości metaheurystyk stosuje się zatem eksperymenty symulacyjne przeprowadzane dla zestawów funkcji testowych (Bartz-Beielstein et al. 2020). Dzięki nim wyniki raportowane przez różnych autorów są porównywalne, łatwiej też uniknąć szeregu dość subtelnych błędów. Tym niemniej, istnieje pole do dalszego rozwoju metodologii benchmarkowania. W referacie przedstawione zostaną główne motywacje przyświecające doborowi problemów testowych oraz kryteriów oceny algorytmów. Następnie omówione będą propozycje dotyczące wprowadzenia nowych planów eksperymentalnych. Idea polega na generowaniu funkcji testowych różniących się interpretowalnymi cechami. Umożliwia to systematyczną, wieloaspektową i interpretowalną ocenę metaheurystyk (Opara et al. 2020).

Literatura

- [1] T. Bartz-Beielstein et al. (2020), *Benchmarking in Optimization: Best Practice and Open Issues*, arXiv preprint arXiv:2007.03488.
- [2] K. Opara et al. (2020), *Parametrized Benchmarking: an outline of the idea and a feasibility study*, Proceedings of the 2020 Genetic and Evolutionary Computation Conference Companion.

Stabilizacja postaci modeli ARIMA dla notowań po debiucie giełdowym

Kacper Paczutkowski
Konrad Furmańczyk

Instytut Informatyki Technicznej SGGW

Rozważamy modele szeregów czasowych klasy ARiMA opisujących notowania spółek giełdowych w ich początkowych okresach od debiutu na Rynku Głównym GPW. Dokonany jest przegląd statystyk ilościowych obrazujących badany obszar rynku GPW w latach jego dotychczasowej działalności. Wyznaczono modele ARiMA dla poszczególnych spółek w każdy z kolejnych dni od ich giełdowego debiutu. Określono także numery notowań, po których dla każdej spółki można było odnotować sekwencje takiej samej klasy modelu. W wyniku badań sformułowano wnioski dotyczące tego ile czasu (notowań) po debiucie musiało upłynąć, by klasa modelu się stabilizowała. Obliczenia wykonano w pakiecie R. Wykorzystano krokowe podejście wyboru modelu za pomocą procedury `auto.arima`, dostępnej w ramach biblioteki *forecast*.

Literatura

- [1] B. Hyndman, Y. Khandakar (2008), *Automatic Time Series Forecasting: The forecast Package for R*, Journal of Statistical Software, Articles, 27(3), 1–22, doi:10.18637/jss.v027.i03

Estymacja przyczynowa pod warunkiem częściowej złej specyfikacji modelu

Krzysztof Rudaś

Wydział Matematyki i Nauk Informatycznych
PW, Instytut Podstaw Informatyki PAN

Modelowanie przyczynowości zajmuje się przewidywaniem efektu podejmowanej przez nas akcji (nowej kampanii marketingowej, metody leczenia pacjentów) dla pojedynczej obserwacji. W tym celu dzielimy populację na grupę eksperymentalną (poddaną działaniu) i kontrolną (nie poddaną działaniu). Następnie konstruujemy estymator który liczy różnicę efektów między grupami dla pojedynczej obserwacji. W [1] pokazaliśmy, przy założeniu liniowości w grupie eksperymentalnej i kontrolnej, asymptotyczne własności dwóch popularnych estymatorów, a także nowego podejścia łączącego zalety dwóch poprzednich.

W moim referacie przedstawię asymptotyczne wyniki uzyskane dla wyżej wymienionych estymatorów przy założeniu, że odpowiedź w grupie kontrolnej i część odpowiedzi w grupie eksperymentalnej, niezwiązana z efektem akcji, są nieliniowe.

Literatura

[1] K. Rudaś, S. Jaroszewicz (2018), *Linear regression for uplift modeling*, Data Mining and Knowledge Discovery 32, 1275–1305

Warunki konieczne i dostateczne sygnaturowej reprezentacji rozkładu czasu życia systemu

Tomasz Rychlik

Instytut Matematyczny PAN

Samaniego (1985) pokazał, że gdy czasy pracy komponentów systemu są niezależne i o jednakowym ciągłym rozkładzie, to rozkład czasu bezawaryjnej pracy systemu jest mieszaniną rozkładów kolejnych awarii komponentów (statystyk pozycyjnych) ze współczynnikami mieszaniny, których wektor zwany sygnaturą systemu, zależy jedynie od struktury systemu. Jest to bardzo użyteczne narzędzie znacznie upraszczające analizę rozkładu czasu pracy systemu. Reprezentacja ta została uogólniona na przypadek permutowalności rozkładu łącznego czasów życia komponentów przez Navarro i in. (2008). Podamy warunki konieczne i dostateczne zachodzenia sygnaturowej reprezentacji Samaniego wyrażone za pomocą ograniczeń na rozkłady brzegowe czasów pracy komponentów i copuli ich zależności. Warunki te pokazują, że reprezentacja Samaniego zachodzi przy znacznie słabszych założeniach niż permutowalność czasów pracy komponentów.

Literatura

- [1] J. Navarro, T. Rychlik, F. Spizzichino (2020), *Conditions on marginals and copula of component lifetimes for signature representation of system lifetime*, w recenzji
- [2] J. Navarro, F.J. Samaniego, N. Balakrishnan, D. Bhattacharya (2008), *Applications and extensions of system signatures in engineering reliability*, *Naval Research Logistics* 55, 313–327
- [3] F.J. Samaniego (1985), *On the IFR closure theorem*, *IEEE Transactions on Reliability*, TR 34, 69–72

Własności asymptotyczne nieparametrycznych estymatorów kwantylowych wersji krzywej Lorenza

Alicja Jokił-Rokita¹, Agnieszka Siedlaczek²

¹Politechnika Wrocławska, Wydział Matematyki

²Uniwersytet Opolski, Wydział Matematyki, Fizyki i Informatyki

Krzywa Lorenza została wprowadzona w 1905 roku przez Lorenza w celu ilustracji koncentracji rozkładu wybranej cechy w populacji. Niestety może być stosowana jedynie w przypadku rozkładów o skończonej wartości oczekiwanej. Prendergast i Staudte (2016) zaproponowali kwantylowe wersje krzywej Lorenza, które mogą być stosowane dla dowolnego rozkładu, a ponadto nie są wrażliwe na obserwacje odstające. W referacie przedstawimy własności asymptotyczne zaproponowanych w pracy Siedlaczek (2018) nieparametrycznych estymatorów kwantylowych wersji krzywej Lorenza.

Literatura

- [1] M. O. Lorenz (1905), *Methods of Measuring the Concentration of Wealth*, Publications of the American Statistical Association 9(70), 209–219
- [2] L. A. Prendergast, R. G. Staudte (2016), *Quantile versions of the Lorenz curve*, Electronic Journal of Statistics 10, 1896–1926
- [3] A. Siedlaczek (2018), *Nonparametric estimation of quantile versions of the Lorenz curve*, Mathematica Applicanda 46(1), 149–157

Estymacja parametrów niejednorodnego procesu gamma z log-liniową funkcją intensywności

Alicja Jokiel-Rokita, Paweł Skoliński

Politechnika Wrocławska, Wydział Matematyki

Niejednorodny proces gamma został zaproponowany przez Bermana (1981). Jest on uogólnieniem procesu Poissona i w praktyce wykorzystywany jest m.in. do modelowania niezawodności systemów naprawialnych. W referacie przedstawimy propozycję modyfikacji estymatorów największej wiarygodności parametrów tego procesu oraz wnioski wynikające z przeprowadzonych symulacji. Przedstawimy również zastosowanie zaproponowanej metody estymacji w analizie danych rzeczywistych.

Literatura

[1] M. Berman (1981), *Inhomogeneous and modulated gamma processes*, *Biometrika*, 68, 143–152

Testy niezależności wielowymiarowych danych funkcjonalnych oparte o brzegową odległość kowariancyjną i brzegowe kryterium Hilberta-Schmidta

Mirosław Krzyśko¹, Łukasz Smaga²

¹Międzywydziałowy Zakład Matematyki i Statystyki, Akademia Kaliska im. Prezydenta Stanisława Wojciechowskiego w Kaliszu

²Wydział Matematyki i Informatyki,
Uniwersytet im. Adama Mickiewicza

Rozważany jest problem testowania niezależności wielowymiarowych danych funkcjonalnych. Korzystając z reprezentacji bazowej danych funkcjonalnych, sprowadzamy ten problem do zagadnienia testowania niezależności danych wektorowych, które mogą być wysoce-wielowymiarowe. W przypadku dwóch procesów, stosujemy testy niezależności oparte o odległość kowariancyjną i kryterium niezależności Hilberta-Schmidta oraz ich brzegowe wersje. W sytuacji wielu procesów, wykorzystane zostały współczynniki wzajemnej zależności agregujące informacje otrzymane z badań par procesów. Rozważane testy niezależności porównane zostały w badaniach symulacyjnych pod względem kontroli prawdopodobieństwa błędu pierwszego rodzaju oraz mocy. Być może najważniejszymi wnioskami z tych badań są: Testy oparte o odległość kowariancyjną i kryterium niezależności Hilberta-Schmidta mają zazwyczaj większą moc niż ich brzegowe wersje, gdy zależność jest liniowa. Natomiast zachodzi sytuacja odwrotna, w przypadku zależności nieliniowej.

Literatura

- [1] A. Gretton, O. Bousquet, A. Smola, B. Schölkopf (2005), *Measuring statistical dependence with Hilbert-Schmidt norms*, In *ALT*, strony 63–77, Springer-Verlag
- [2] A. Gretton, K. Fukumizu, C.-H. Teo, L. Song, B. Schölkopf, A. Smola

(2008), *A kernel statistical test of independence*, In *Advances in Neural Information Processing Systems* 20, strony 585–592, Cambridge, MA. MIT Press

[3] Z. Jin, D.S. Matteson (2018), *Generalizing distance covariance to measure and test multivariate mutual dependence via complete and incomplete V-statistics*, *Journal of Multivariate Analysis* 168, 304–322

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

[5] G.J. Székely, M.L. Rizzo, N.K. Bakirov (2007), *Measuring and testing dependence by correlation of distances*, *The Annals of Statistics* 35, 2769–2794

[6] G.J. Székely, M.L. Rizzo (2013), *The distance correlation t-test of independence in high dimension*, *Journal of Multivariate Analysis* 117, 193–213

[7] C. Zhu, X. Zhang, S. Yao, X. Shao (2020), *Distance-based and RKHS-based dependence metrics in high dimension*, *The Annals of Statistics* (przyjęta do druku)

Normal distribution with plasticizing component

Piotr Sulewski sdfs

Pomeranian University, Institute of Exact and Technical Sciences
Arciszewskiego Street 22, 76-200 Slupsk, Poland,
ORCID ID: 0000-0002-0788-6567, e-mail: piotr.sulewski@apsl.edu.pl

In this paper we introduce a new flexible plasticizing component and mixture of two distributions, further called the normal distribution with the plasticizing component. Some properties of the introduced distributions are derived, such are the cumulative distribution and the probability density function, moments, skewness and kurtosis. The unknown parameters are estimated by the maximum likelihood method. Illustrative examples of applicability and flexibility of the introduced distributions are given. We compare the flexibility of the introduced distribution and various distributions using skewness and kurtosis. asdad

Literatura

- [1] A. Azzalini (1985), *A class of distributions which includes the normal ones*, Scandinavian Journal of Statistics, 171–178
- [2] A. Behboodian (1970), *On the Modes of a Mixture of Two Normal Distributions*, Technometrics 12(1), 131–139
- [3] H. Bolfarine, G. Martínez-Flórez, H.S. Salinas (2018), *Bimodal symmetric-asymmetric power-normal families*, Communications in Statistics-Theory and Methods 47(2), 259–276
- [4] P. Sulewski (2019), *Two-Piece Power Normal Distribution*, Communications in Statistics-Theory and Methods DOI:10.1080/03610926.2019.1674871
- [5] I. Yadegari, A. Gerami, M.J. Khaledi (2008), *A generalization of the Balakrishnan skew-normal distribution*, Statistics and Probability Letters 78, 1165–1167.

Wpływ sygnatur na właściwości rozkładu czasu życia systemu w klasycznym modelu wykładniczym iid.

Magdalena Szymkowiak

Politechnika Poznańska

W referacie omówiony zostanie wpływ sygnatur na właściwości rozkładu czasu życia systemu zbudowanego z identycznych niezależnie działających elementów o wykładniczych czasach życia. W szczególności zaprezentujemy wyniki związane z silną jednomodalnością rozkładu czasu życia systemu a także monotonicznością intensywności uszkodzeń oraz średniego pozostałego czasu życia systemu. Przedstawione wyniki zostały otrzymane we współpracy z Tomaszem Rychlikiem z IM PAN.

Literatura

- [1] Navarro, J., Rubio, R. (2009), *Computations of signatures of coherent systems with five components*, Communications in Statistics – Simulation and Computation, 39, 68–84
- [2] Rychlik, T., Szymkowiak, M. (2020), *Signature conditions for distributional properties of system lifetimes if component lifetimes are iid exponential*, submitted
- [3] Samaniego, F.J. (1985), *On closure of the IFR class under formation of coherent systems*, IEEE Transactions on Reliability, R-34, 69–72

Klasyfikacja wieloetykietowa z niepełną obserwowalnością etykiet

Paweł Teisseyre

Instytut Podstaw Informatyki PAN oraz Wydział
Matematyki i Nauk Informacyjnych PW

Przedstawię problem klasyfikacji wieloetykietowej z niepełną obserwowalnością etykiet. W tradycyjnym problemie klasyfikacji wieloetykietowej zakłada się że zbiór uczący zawiera poprawnie przypisane etykiety dla wszystkich obserwacji. To założenie może być w wielu sytuacjach niespełnione, np. w przypadku prognozowania wielozachorowalności (współwystępowania wielu chorób u jednego pacjenta) może się zdarzyć że pewne choroby nie zostały zdiagnozowane i nie ma wówczas pewności że występują lub nie u danego pacjenta. W rozważanym schemacie, przyjmujemy że jeśli etykieta została przypisana to na pewno jest związana z daną obserwacją, zaś brak przypisania może oznaczać zarówno to że jest ona związana z daną obserwacją jak i to że nie jest z nią związana. Omówię podstawowe własności teoretyczne takiego schematu. Pokażę w jaki sposób zmodyfikować łańcuchy klasyfikatorów (będące najpopularniejszym algorytmem w klasyfikacji wieloetykietowej) do powyższego schematu.

Literatura

- [1] J. Bekker, J. Davis (2020), *Learning from positive and unlabeled data: a survey*, Machine Learning 109 (4), 719-760
- [2] P. Teisseyre (2020), *Classifier chains for positive unlabelled multi-label learning*, under review

Asymptotyka optymalnego obszaru ufności dla parametrów położenia i skali na podstawie statystyk porządkowych: losowy rozmiar próby

Aleksander Zaigrajew

Uniwersytet Mikołaja Kopernika w Toruniu

Rozważamy zagadnienie konstrukcji dwuwymiarowego optymalnego obszaru ufności, czyli obszaru ufności o najmniejszym polu, dla parametrów położenia i skali na podstawie statystyk porządkowych z próby x_1, \dots, x_n . W przypadku nielosowego rozmiaru próby i dwóch statystyk porządkowych asymptotyka takiego obszaru, gdy rozmiar próby $n \rightarrow \infty$, była otrzymana w [3].

Przy pomocy wyników, otrzymanych w [1,2], wynik ten zostanie poszerzony na przypadek $k \geq 2$ statystyk porządkowych i próby losowego rozmiaru $N(n)$ przy następującym założeniu: $(N_n)_{n \geq 1}$ jest ciągiem całkowitoliczbowych nieujemnych zmiennych losowych, takich że N_n i x_1, \dots, x_n, \dots są niezależne dla każdego n oraz $N_n \rightarrow \infty$ według prawdopodobieństwa, a $N_n/n \implies Y$ według rozkładu, gdy $n \rightarrow \infty$, gdzie Y jest pewną zmienną losową o rozkładzie absolutnie ciągłym.

Literatura

- [1] V. E. Bening, V. Yu. Korolev (2005), *On an application of the Student distribution in the Theory of Probability and Mathematical Statistics*, Theory Probab. Appl. 49, 377–391
- [2] V. Yu. Korolev (2000), *Asymptotic properties of sample quantiles constructed from samples with random sizes*, Theory Probab. Appl. 44, 394–399
- [3] A. Zaigraev, M. Alama-Bučko (2006), *Asymptotics of the optimal confidence region for shift and scale, based on two order statistics*, Statistical Methods of Estimation and Testing Hypotheses, Perm University, 49–65. Translated in: J. Math. Sci. 2017, 220, 6, 763–776