

ROBUST 2008



Račkova dolina
hotel Mier

ROBUST 2008

Definitní borník abstraktů pro web

Antoch Jaromír, Mielle Rafaele	
<i>A note on variability of interval data</i>	9
Arendacká Barbora	
<i>Aproximácia zovšeobecného konfidenčného intervalu pre σ_1^2</i>	9
Boháčová Hana	
<i>Oblasti necitlivosti pro varianční komponenty ve smíšeném lineárním regresním modelu</i>	9
Branda Martin	
<i>Úlohy s pravděpodobnostními omezeními a jejich reformulace</i> ...	10
Branda Martin	
<i>Vývoj uživatelsky přátelského statistického softwaru</i>	11
Bubelíny Peter	
<i>Podiely génových expresií v analýze microarray dát</i>	11
Cavaliere Roberto	
<i>Mathematica 6: Data analysis, visualization and interaction</i>	11
Cézová Eliška	
<i>Zónové regulační diagramy</i>	12
Cimermanová Katarína	
<i>Klasifikácia zašumených dát</i>	13
Černíková Alena, Hušková Marie, Prášková Zuzana, Steinebach Josef	
<i>Zpoždění odhadu bodu změny při sekvenčním přístupu</i>	14
Čížek Pavel	
<i>Robust estimation of instrumental variable regression</i>	14
Desatníková Júlia	
<i>Metóda maximálnej empirickej vierohodnosti v časových radoch</i>	15
Dohnal Gejza	
<i>Design of control charts</i>	15
Dokoupilová Pavla	
<i>Ověřování účinnosti léčby (simulační studie)</i>	16
Dostál Petr	
<i>Robust filtering and portfolio management</i>	16

Došlá Šárka	
<i>Time series with non-positive autocorrelations</i>	17
Fabián Zdeněk	
<i>O rozděleních s těžkými chvosty</i>	17
Fajfrová Lucie	
<i>Model dopravní špičky z mikroskopického pohledu</i>	17
Friessl Michal	
<i>Koziolův-Greenův model se zleva useknutými pozorováními</i>	18
Grendár Marian	
<i>Podmienený zákon veľkých čísel</i>	19
Hančová Martina	
<i>Odhady variančních parametrů v LRM s konečným diskretním spektrem</i>	19
Hanzák Tomáš	
<i>Exponenciální vyrovňávání pro nepravidelné časové řady</i>	19
Harman Radoslav	
<i>Algoritmy optimálneho navrhovania experimentov</i>	20
Helisová Kateřina	
<i>Model pro náhodné sjednocení kruhů se vzájemnými interakcemi</i>	20
Hlávka Zdeněk	
<i>Funkcionální podmíněné kvantily rizikově neutrálních hustot</i>	21
Hlubinka Daniel, Kotík Lukáš, Vencálek Ondřej	
<i>Výprava do hlubin dat</i>	21
Hönschová Erika	
<i>Estimating the scale parameter in Burr distribution</i>	22
Hornišová Klára	
<i>Rekurentné vzorce pre počty niektorých druhov permutácií</i>	22
Hušková Marie, Horváth Lajos, Kokoszka Piotr	
<i>Testing the stability of the functional autoregressive process</i>	23
Chochola Ondřej	
<i>Kvantily v sekvenční analýze bodu změny</i>	23
Janáček Jiří	
<i>Řešení problémů analýzy obrazu pomocí MAP a minima totální variace</i>	23
Janžura Martin	
<i>O významu entropie</i>	24
Jarušková Daniela	
<i>Testování normality ve vícerozměrném případě</i>	25
Jirkovský Jaroslav	
<i>Využití MATLABu ve statistice a optimalizaci</i>	25

Jonáš Petr	
<i>Vektorové autoregresní modely</i>	26
Jurczyk Tomáš	
<i>Robustifikace hřebenové regrese – první pokusy</i>	26
Kalousová Anna	
<i>Joseph-Émile Barbier a stereologie v XIX. století</i>	27
Karlová Andrea	
<i>Levy processes and balayage theory</i>	27
Klaschka Jan	
<i>Intervalový odhad parametru p binomického rozdělení:</i>	
<i>Co je (relativně) nového?</i>	28
Kláštorecký Petr, Kulich Michal	
<i>Combined logistic regression estimator for the case-cohort</i>	
<i>proportional odds model</i>	29
Komárek Arnošt	
<i>R balíček pro odhad jedno i vícerozměrných rozdělení pomocí</i>	
<i>normálních směsí, případně s intervalově cenzorovanými daty</i> ...	30
Kotík Lukáš	
<i>Směrové kvantily pro vícerozměrná data</i>	30
Kovářová Milena	
<i>Text mining of articles in metabolomics research</i>	30
Král Pavol, Lašová Lenka, Ďurica Marek	
<i>Základy statistiky s fuzzy datami</i>	31
Kvaszová Milena	
<i>Proč nám nerozumějí?</i>	31
Kvitkovičová Andrea	
<i>Testy hypotéz o parametrech reškálovaného Wienerovho</i>	
<i>procesu s posunutím</i>	32
Lachout Petr	
<i>Model pro předávání genů</i>	32
Lechnerová Radka, Lechner Tomáš	
<i>Analýza rozmístění elektronických podatelen obcí v České</i>	
<i>republice</i>	32
Linka Aleš, Tunák Maroš, Bajzík Vladimír	
<i>Detekce defektů ve tkanině pomocí vícerozměrných</i>	
<i>regulačních diagramů</i>	33
Maciak Matúš	
<i>Spline models with change points applied to seroprevalence</i>	
<i>data</i>	33

Madurkayová Barbora	
<i>Ratio type statistics for detection of changes in linear regression models</i>	34
Marek Jaroslav	
<i>Příběh z cyklu „Statistický pohled na měření v české a světové literatuře“</i>	34
Marek Tomáš, Tichavský Petr, Darbellay Georges, Franěk Jaroslav	
<i>O odhadu vzájemné informace</i>	35
Marušiaková Miriam	
<i>Multiple changes in coefficients of autoregressive models</i>	35
Michálek Jiří	
<i>Robustní statistické návrhy</i>	36
Neubauer Jiří, Veselý Vítězslav	
<i>Change point detection by basis pursuit</i>	37
Orava Jan	
<i>Porovnání metod odhadu vyhlazovacího parametru při jádrových odhadech hustoty</i>	38
Pazdera Jaroslav	
<i>Detecting atoms in deconvolution</i>	38
Pázman Andrej	
<i>Ako lepšie porozumieť nelineárnej metóde najmenších štvorcov pomocou geometrie</i>	38
Pečánka Jakub	
<i>Change detection in autoregressive time series</i>	39
Pelikán Emil, Brabec Marek, Malý Marek, Konár Ondřej	
<i>Zkušenosti se statistickým modelováním spotřeby zemního plynu</i>	40
Pešta Michal	
<i>Regression in Sobolev spaces using total least squares</i>	40
Pícek Jan	
<i>Poznámky k metodě POT</i>	41
Plát Pavel	
<i>Nejmenší vážené čtverce v příkladech</i>	41
Prokešová Michaela	
<i>Moment estimates for spatial cluster point processes</i>	41
Půlpán Zdeněk, Čihák Michal, Bomberová Jaroslava	
<i>Fuzzy míry neurčitosti měření za nejistoty a v podmínkách vágnosti</i>	42
Ročková Veronika	
<i>Moderní přístupy k testování periodicity v časových řadách</i>	42

Rosipal Roman	
<i>Probabilistic modelling of the sleep process</i>	43
Ryšavá Eva	
<i>Determinanty PZI ve zpracovatelském průmyslu ČR v letech 2000–2006</i>	43
Řezanková Hana, Húsek Dušan	
<i>Zjišťování optimálního počtu shluků ve statistických programových systémech</i>	44
Saxl Ivan	
<i>Pravděpodobnost ve středověku</i>	45
Shokirov Bobosharif	
<i>Hypothesis testing by using N-distances</i>	46
Schindler Martin	
<i>Rank score test for independence with nuisance nonlinear regression</i>	46
Somorčík Ján	
<i>Viacrozmerné testy o parametroch polohy viacerých súborov</i>	46
Staněk Jakub	
<i>Stochastický model vývoje epidemie s vakcinací</i>	47
Strouhal Jan	
<i>Jak posuzovat spolehlivost softwaru</i>	47
Šedová Michaela, Kulich Michal	
<i>Maximálně věrohodné odhady a lineární regrese ve výběrových šetřeních</i>	48
Šimeček Petr	
<i>Problém s příliš velkými daty – animal model</i>	48
Šimečková Marie	
<i>Testy aditivity v analýze dvojného třídění bez opakování</i>	48
Štěpán Josef	
<i>Mathematical models for epidemics</i>	49
Timková Jana	
<i>Bayesian nonparametric estimation of hazard rate in survival analysis using Gibbs sampler</i>	49
Tuček Pavel, Tuček Jiří, Marek Jaroslav, Procházka Roman	
<i>Nelineární statistická inference v oblasti nanotechnologií</i>	50
Tvrdík Josef	
<i>Adaptivní stochastické algoritmy v nelineární regresi</i>	50
Václavík Vladimír	
<i>Metoda MLE pro logaritmicko-normální rozdělení s prahem</i>	51

Vajda Igor	
<i>On efficiencies of decisions about statistical models based on f-divergences of empirical distributions</i>	51
Vencálek Ondřej	
<i>Některé vlastnosti váhové poloprostorové hloubky</i>	52
Víšek Jan Ámos	
<i>Consistency of weighted generalized method of moments</i>	52
Volf Petr, Linka Aleš	
<i>Odhadování nerovnoměrnosti zatížení v paralelním systému</i>	54
Vokáčová Kateřina, Marek Patrice, Pavelka Tomáš, Neumanová Martina, Šedivá Blanka, Vávra František	
<i>Ratio statistics – tolerance intervals</i>	54
Wimmer Gejza	
<i>Přibližné konfidenční intervaly pre meraní hodnotu v prípade digitalizovaných meraní</i>	54
Wimmer Gejza Jr.	
<i>Metódy výpočtu približných konfidenčných intervalov parametra polohy z digitalizovaných meraní</i>	55
Witkovský Viktor	
<i>Konfidenčné intervaly založené na digitalizovaných meraniach</i>	55
Žambochová Marta	
<i>Shlukování velkých souborů dat pomocí metod rozkladu</i>	56
ČStS	
<i>Staň se členem naší společnosti – ještě dnes!</i>	57
Generali a.s.	
<i>Nabídka zaměstnání</i>	58
Humusoft spol. s r.o.	
<i>Cena pro studenty</i>	59
Elkan spol. s r.o.	
<i>Cena pro studenty</i>	60

Naber smer hotel Mier



Račkova dolina
5 km

A note on variability of interval data

Jaromír Antoch, Raffaele Miele

MFF UK, KPMS, Sokolovská 83, CZ–186 75 Praha 8

jaromir.antoch@mff.cuni.cz

The aim of this lecture is two-fold. First, when we have interval valued observations, what is the range of possible values of their variance. In parallel, we respond to the same question when the variance is replaced by the mean of the absolute values of the residuals from the median. Second, to show that it is very important to take care how can “differ” potential data from these intervals when variance, mean of the absolute values of the residuals from the median respectively, is taken as the measure of their difference. Several solutions will be discussed.

Aproximácia zovšeobecneného konfidenčného intervalu pre σ_1^2

Barbora Arendacká

ÚM SAV, Dúbravská cesta 9, SK–841 04 Bratislava

barendacka@gmail.com

V príspevku sa budeme zaoberať intervalovými odhadmi pre variančný komponent zodpovedajúci náhodnému efektu v zmiešaných lineárnych modeloch s dvomi variančnými komponentami. Konkrétne pôjde o zovšeobecnený konfidenčný interval založený na zovšeobecnenom pivote, ktorý navrhli Park a Burdick (Communications in Statistics, Simul. and Comp., 32, 2003), a približný, El-Bassiouniho-Williamsov-Tukeyho interval (El-Bassiouni, Communications in Statistics, Theory and Methods, 23,1994). Na základe doterajších simulačných výsledkov a formúl pre výpočet sa obe procedúry zdajú byť veľmi príbuzné. V príspevku na tieto podobnosti upozorníme a zároveň ukážeme, že El-Bassiouniho-Williamsov-Tukeyho interval môžeme v istom zmysle považovať za aproximáciu spomenutého zovšeobecneného intervalu. Tento výsledok sa odvíja od postupu, ktorý na jednoduchšiu verziu uvažovanej situácie aplikoval Bross (Biometrics 6, 1950).

Oblasti necitlivosti pro varianční komponenty ve smíšeném lineárním regresním modelu

Hana Boháčová

UPCE, FES, Ústav matematiky, Studentská 84, CZ–532 10 Pardubice

hana.bohacova@upce.cz

K určení maximálně věrohodných odhadů parametrů střední hodnoty β i variančních komponent θ ve smíšeném lineárním regresním modelu je třeba použít nějaké počáteční hodnoty variančních komponent jako vstup do iterační procedury. Otázkou samozřejmě je, jak tyto počáteční hodnoty volit a jak se změní odhady, pokud použijeme počáteční hodnoty jiné. Zdá se být pochopitelné požadovat, aby při změně počátečních hodnot variančních komponent nedošlo k přílišnému nárůstu disperzí jednotlivých složek odhadu. Zvolíme-li nějaké pevně dané počáteční hodnoty variančních komponent θ_0 a vektor \mathbf{g} , který bude určovat lineární kombinaci složek odhadu θ , případně β , můžeme pak najít množinu všech počátečních hodnot $\theta_0 + \delta\theta$, jejichž použití nezpůsobí nárůst disperze odhadu $\mathbf{g}'\theta$ (respektive $\mathbf{g}'\beta$) o více než o ε -násobek oproti odhadu získanému pomocí počáteční hodnoty θ_0 , pro předem zvolené malé kladné ε . Takovou množinu budeme nazývat oblastí necitlivosti pro varianční komponenty, respektive oblastí necitlivosti pro parametry střední hodnoty. V obou případech - pro varianční komponenty i pro parametry střední hodnoty - lze oblast necitlivosti určit pomocí kvadratické formy. Je snadné ukázat, že v případě oblasti necitlivosti pro parametry střední hodnoty je matice této kvadratické formy singulární a že vektor θ_0 je kolmý na podprostor generovaný sloupci této matice. Pro oblast necitlivosti pro varianční komponenty je matice určující příslušnou kvadratickou formu poměrně komplikovaná a její singularitu se zatím obecně prokázat nepodařilo. Na základě několika zpracovaných numerických studií se však ukazuje, že i tato matice singulární je. Otázkou zůstává, zda je tato singularita náhodná nebo zda i v případě oblasti necitlivosti pro varianční komponenty má matice příslušné kvadratické formy obecně podobné vlastnosti jako v případě paramterů střední hodnoty.

Úlohy s pravděpodobnostními omezeními a jejich reformulace

Martin Branda

MFF UK, KPMS, Sokolovská 83, CZ-186 75 Praha 8

branda@karlin.mff.cuni.cz

Příspěvek se zabývá možností přeformulovat úlohu s pravděpodobnostními omezeními pomocí úlohy stochastického programování s penalizací v účelové funkci. Takovéto úlohy jsou poté snadněji řešitelné a i jejich struktura umožňuje snažší studium jejich vlastností. Je odvozena asymptotická ekvivalence obou úloh.

Vývoj uživatelsky přátelského statistického softwaru

Martin Branda

Median s.r.o. a Lerach s.r.o.

branda@karlin.mff.cuni.cz

V příspěvku bude stručně představena firma MEDIAN, jejíž softwarová část se zabývá vývojem softwaru pro zpracování dat a mediální plánování. Podrobněji se zaměříme na Data Analyzer, kde jsou uživatelsky přátelským způsobem implementovány moderní statistické metody jako například faktorová a shluková analýza, vícerozměrné škálování, korespondenční analýza a další. K nejzajímavějším metodám jistě patří klasifikační a regresní stromy (MCART), které implementují rozšíření klasické metody CART s nebinárními štěpeními. Formou kvízu bude též představena unikátní sbírka testů rovnosti středních hodnot.

Podiely génových expresií v analýze microarray dát

Peter Bublín

MFF UK, KPMS, Sokolovská 83, CZ-186 75 Praha 8

bubeliny@karlin.mff.cuni.cz

Jeden z cieľov microarray experimentov je nájsť odlišne expresované gény získané z dvoch alebo viacerých štádií chorôb. Problém je, že génové expresie sú vysoko korelované. V tejto práci sú uvažované podiely génových expresií vytvorených z neusporiadaných alebo z usporiadaných párov génov. Pre HYPERDIP a TEL dáta (odlišné štádiá detskej leukémie) sa zdá, že tieto podiely pre rôzne páry sú približne nezávislé. Pre každú situáciu sú odhadnuté p-hodnoty pre testovanie, ktoré gény alebo ich podiely sú odlišne expresované. Tieto odhadnuté p-hodnoty sú zobrazené v histogramoch a tvary týchto histogramov sú porovnané. Porovnávacia štúdia ukázala, že tvary histogramov pre p-hodnoty spočítané z génových expresií a z ich podielov sa výrazne líšia.

Mathematica 6: Data analysis, visualization and interaction

Roberto Cavaliere

Wolfram Research

roberto@wolfram.com

Mathematica's capabilities in statistics-related areas have steadily grown. It includes industrial-strength versions of the usual statistics and visualization

capabilities, together with many important unique capabilities, made possible by its broad overall scope and extensibility, immense interconnected web of algorithms and integrated programming language and interface framework.

Mathematica provides integrated support both for classical statistics and for modern large-scale data analysis. Incorporating the latest numeric and computational geometry algorithms, Mathematica provides high-accuracy and high-reliability statistical results for datasets of almost unlimited size.

Beyond the internal algorithms and data analysis functionalities, Mathematica provides additional facilities useful in typical data analysis processes. For instance, it imports and exports data in more than one hundred standard formats, allowing the data exchange with other applications. It has a built-in high-level interface to all standard SQL databases, that allows immediate searching, reading and writing of arbitrary data and expressions. It gives access to a library of carefully cured and continually updated data maintained by Wolfram Research; for instance Financial Data, Chemical Data, Country and City Data and lots more.

Finally, Mathematica automates the generation of compelling visualizations – and new technology in Mathematica 6 allows any output or computation to immediately become interactive and dynamic. Its integrated notebook mechanism, enable to create documents which can store a complete, formatted, editable, executable history of an analysis, complete with live graphics, interactive controls and arbitrary typeset tabular, mathematical and other material.

This presentation will show some examples about the aforementioned features related to data analysis processes.

Zónové regulační diagramy

Eliška Cézová

ČVUT, CQR FS, Karlovo nám. 13, CZ–121 35 Praha 2

eliskacqr@email.cz

Zónový regulační diagram je speciálním typem regulačního diagramu pro \bar{X} a R . Zónový regulační diagram je alternativou k Shewhartovu regulačnímu diagramu, respektující určitá „doplňková pravidla“, tzv. *supplementary run rules*. K vyhodnocení stability procesu používá kromě informace získané z aktuálního výběru i informaci o předchozích měřeních. Tím se podobá diskrétní verzi CUSUM diagramu, tzv. *Run sum chart*, ale je ke mnohem jednodušší jak pro porozumění tak i pro implementaci. Jednoduchost aplikace jej předurčuje k automatické regulaci.

V příspěvku je uveden nový návrh flexibilního zónového diagramu pro \bar{X} , který odstraňuje nevýhody klasického návrhu a svými vlastnostmi se blíží ke složitějšímu CUSUM diagramu. Zároveň je prezentována metoda pro výpočet průměrné doby běhu (ARL) tohoto diagramu pomocí Markovského řetězce. V tabulce jsou srovnávány různé varianty zónového diagramu pomocí ARL při různých hodnotách posunu střední hodnoty procesu oproti předpokládané (stabilní) nulové hodnotě. Ve spojení s ekonomicko-statistickou optimalizací parametrů může tento diagram představovat velmi dobrou alternativu pro automatickou regulaci procesu.

Klasifikácia zašumených dát

Katarína Cimermanová

ÚM SAV, Dúbravská cesta 9, SK – 841 04 Bratislava

katarina.cimermanova@gmail.com

Klasifikácia viacrozmerných pozorovaní do dvoch tried je dôležitý problém. Existuje niekoľko klasifikačných metód na zatriedenie pozorovaných vektorov do jednej z dvoch tried, avšak v reálnom živote sú vektory pozorovaní zašumené. Riešením klasifikácie zašumených dát je robustná formulácia vychádzajúca z metódy oporných bodov. Formulácia je konvexný optimalizačný problém, ktorý je súčasťou problematiky kónického programovania druhého rádu. V robustnej formulácii sa predpokladá elipsoidálny model šumu. Nie je nutný predpoklad typu rozdelenia pozorovaných dát, predpokladá sa len konečnosť momentov druhého rádu.

Robustnú klasifikačnú metódu aplikujeme na analýzu vydychovaných plynov, kde sa budeme venovať klasifikácii dobrovoľníkov do skupiny fajčiarov a nefajčiarov.

Reference

- [1] Bi J. and Zhang T. (2004) *Support Vector Classification with Input Data Uncertainty*. Advances in Neural Information Processing systems (NIPS'4) **17**, 161–168.
- [2] Sturm J.F. (1995) *Using SEDUMI 1.02, a Matlab*toolbox for Optimization over symmetric cones*. (Updated for Version 1.05), Optimization Methods and Software **11**, 625–653.
- [3] Bhattacharyya Ch. (2004) *Robust Classification of noisy data using Second Order Cone Programming approach*. In Proceedings International Conference on Intelligent Sensing and Information Processing, 433–438.
- [4] Shivaswamy P.K., Bhattacharyya Ch. and Smola A.J. (2006) *Second Order Cone Programming Approach for Handling Missing and Uncertain Data*. Journal of Machine Learning Research **7**, 1283–1314.

Zpoždění odhadu bodu změny při sekvenčním přístupu

Černíková Alena, Marie Hušková, Zuzana Prášková, Josef Steinebach

KPMS MFF UK, Sokolovská 83, CZ – 186 75 Praha 8;

Universität zu Köln, Mathematisches Institut, Germany;

ÚAMVT PŘF UK, Albertov 6, CZ – 120 00 Praha 2

aja@matfyz.cz, marie.huskova@mff.cuni.cz, zuzana.praskova@mmf.cuni.cz

Příspěvek se zabývá analýzou bodu změny v postupně přicházejících datech. Předpokládáme data přicházející jednotlivě, která splňují model lineární regrese

$$Y_i = \beta_i^T \mathbf{X}_i + e_i, i = 1, 2, \dots$$

kde Y_i jsou pozorovaná data, β_i jsou regresní koeficienty, \mathbf{X}_i jsou vektory nezávislých proměnných a e_i jsou náhodné chyby. K dispozici jsou tzv. tréninková data o rozsahu m , která jsou beze změny. Cílem je potom odhalit případnou změnu v regresních koeficientech β_i , která může nastat v libovolném, předem neznámém čase $m + k^*$, a přitom dodržet předem dané požadavky na pravděpodobnost falešného poplachu a na pravděpodobnost odhalení skutečné změny.

Postup pro detekci bodu změny v takovémto případě vychází z teorie testování hypotéz a čas detekce $\tau(m)$ je definován jako

$$\tau(m) = \inf \{k \geq 1 : Q(m, k) \geq cq^2(m, k, \gamma)\},$$

kde $Q(m, k)$ je detektor založený na kumulativní ch součtech residuí $\hat{e}_i = Y_i - \hat{\beta}_m^T \mathbf{X}_i$, $i = m + 1, \dots, m + k$ ($\hat{\beta}_m$ je odhad parametru β z tréninkových dat), c je kritická hodnota, určená na základě omezení na pravděpodobnost falešného poplachu, a $q(m, k, \gamma)$ je tzv. hraniční funkce závislá navíc na konstantě γ .

Při určitých předpokladech na parametry modelu, okamžik změny k^* a velikost změny $\Delta_m = |\beta_0 - \beta_*|$, kde β_0 je hodnota regresního koeficientu před změnou a β_* je jeho hodnota po změně, lze dokázat, že čas detekce $\tau(m)$ má asymptoticky normální rozdělení pro $m \rightarrow \infty$.

Příspěvek je zaměřen na ukázkou rychlosti konvergence k limitnímu normálnímu rozdělení pro různé hodnoty parametrů modelu. Výsledky jsou získány prostřednictvím simulací v programu R.

Robust estimation of instrumental variable regression

Pavel Čížek

Tilburg University, Fac. of Economics and Business Administration, Room K616,
P.O.Box 90153, NL – 5000LE Tilburg, The Netherlands

p.cizek@uvt.nl

Instrumental variable regression is a technique for estimating regression relationship with endogenous regressors. Being typically estimated by the two-stage least squares

(2SLS), the procedure is very sensitive to outlying observations and extreme values in both explanatory variables and instruments. To reduce this sensitivity of 2SLS, procedures based on the L_1 - and M -estimators of regression were proposed, for example, by [1], [2] and [3], [4], respectively. These methods are however not robust to outliers with large leverage (e.g., [1]) or large values of instruments (e.g., [2]) or restrict the class of error distributions ([4]).

Here we propose several ways for estimating the regression models with endogenous regressors. On one hand, we demonstrate a simple procedure to obtain positive breakdown-point quantile regression by means of data transformations. On the other hand, we propose a generalized-method-of-moments (GMM) estimator based on trimmed moment conditions, which allows for high breakdown-point estimation of regression parameters under weak identification assumptions.

Reference

- [1] Chernozhukov V. and Hansen C. (2008) *Instrumental variable quantile regression: a robust inference approach*. Journal of Econometrics **142**, 379–398.
- [2] Honore B. E. and Hu L. (2004) *On the Performance of Some Robust Instrumental Variables Estimators*. Journal of Business & Economic Statistics **22(1)**, 30–39.
- [3] Ronchetti E. and Trojani F. (2001) *Robust inference with GMM estimators*. Journal of Econometrics **101**, 37–69.
- [4] Wagenvoort R. and Waldman R. (2002) *On B-robust instrumental variable estimation of the linear model with panel data*. Journal of Econometrics **106**, 297–324.

Metóda maximálnej empirickej vierohodnosti v časových radoch

Júlia Desatníková

FPV UMB, KM, Tajovského 40, SK–974 01 Banská Bystrica

desatnik@fpv.umb.sk

V príspevku sa budeme venovať metóde maximálnej empirickej vierohodnosti, jej popisu a využitiu v analýze časových radov na robustné odhadovanie parametrov. Toto využitie si prakticky ukážeme na AR(1) procese. Výsledky porovnáme s niektorými publikovanými výsledkami.

Design of control charts

Gejza Dohnal

ČVUT, CQR FS, Karlovo nám. 13, CZ–121 35 Praha 2

dohnal@nipax.cz

Regulační diagramy uvedl Dr. Shewhart na počátku třicátých let minulého století jako nástroj pro sledování jakostních znaků při hromadné výrobě. Ukazuje se, že

ač se jimi od té doby zabývala řada významných matematiků a inženýrů, v praxi se stále používají téměř v nezměněné podobě. Kolem regulačních diagramů se vede mnoho diskusí souvisejících jak s jejich teoretickým zázemím tak i se způsobem aplikace.

Při aplikaci statistické regulace výrobního procesu prostřednictvím regulačních diagramů je třeba postupovat obezřetně, na základě promyšleného návrhu. Příspěvek má za cíl ukázat hlavní kroky při návrhu regulačního diagramu, mezi něž patří vedle důkladné analýzy výrobního procesu a jeho pravděpodobnostních charakteristik také výběr vhodného typu regulačního diagramu a optimalizace jeho parametrů. Ekonomicko-statistický návrh bere do úvahy jak ekonomické parametry (náklady na měření, identifikaci vymezitelné poruchy, zásah v případě falešného poplachu a v případě oprávněné signalizace, ztráty z provozu mimo kontrolu a další), tak i statistické vlastnosti (pravděpodobnost vzniku poruchy, předpokládanou dobu, po kterou proces bude pod kontrolou, předpokládanou dobu běhu mimo kontrolu, pravděpodobnost falešného poplachu a další). V případě, kdy se jedná o proces obnovy, je sledována střední cena (náklad) na hodinu provozu, která je spočtena jako podíl středního nákladu na cyklus a střední délky cyklu. Optimalizace se provádí vzhledem ke ztrátové funkci, která vychází z Tagucioho ztrátové funkce.

Ukazuje se, že pro srovnání různých typů regulačních diagramů stačí porovnat jejich střední délku běhu (Average Run Length, ARL) při různých posuvech střední hodnoty sledované charakteristiky.

Ověřování účinnosti léčby (simulační studie)

Pavla Dokoupilová

pavla@dokoupilova.cz

Základní nástrojem pro měření rozdílu účinností mezi dvěma typy léčby je v případě binárních dat poměr šancí. Studie se zabývá vlastnostmi dvou základních odhadů používaných pro poměr šancí a je prováděna pomocí simulace klinických pokusů s různým počtem pacientů v léčené i v kontrolní skupině a s různými pravděpodobnostmi úspěchu. Cílem příspěvku je poskytnout přehled o základních aspektech chování odhadů.

Robust filtering and portfolio management

Petr Dostál

MFF UK, KPMS, Sokolovská 83, CZ – 186 75 Praha 8

petr.dostal@mff.cuni.cz

We consider a non-consuming agent investing in a stock and a money market interested in the portfolio market price far in the future. We derive an almost log-optimal strategy for small proportional transaction costs in a model with stochastic coefficients. If the rate of return cannot be observed directly, we use robust filtering in order to obtain a model with observable stochastic coefficients. We also modify the

obtained strategy for an investor who is more risk averse so that the strategy almost maximizes the asymptotics of expected HARA utility in case of deterministic coefficients.

Time series with non-positive autocorrelations

Šárka Došlá

MFF UK, KPMS, Sokolovská 83, CZ-186 75 Praha 8

`dosla@karlin.mff.cuni.cz`

We deal with time series models with non-positive autocorrelations. For a strictly stationary process $\{X_t, t \in \mathbb{Z}\}$ with an autocorrelation function $\{r_t\}_{t=1}^{\infty}$ such that $r_t \leq 0$ for all $t \geq 1$ the lower boundary for the sum $\sum_{t=1}^{\infty} r_t$ is investigated. Some results are introduced for general stationary random processes, but the main part of this work is devoted to dependent binary variables. The case of 1-dependent variables is considered and the boundaries for the 1 lag autocorrelation r_1 are investigated using some special models. Properties of the sum $\sum_{t=1}^{\infty} r_t$ in a more general situation are studied as well. We show that for Bernoulli variables obtained by clipping a Gaussian process the autocorrelation function $\{r_t\}_{t=1}^{\infty}$ with $r_t \leq 0$ for all $t \geq 1$ satisfies $\sum_{t=1}^{\infty} r_t \geq -1/3$.

O rozděleních s těžkými chvosty

Zdeněk Fabián

ÚI AV ČR, Pod Vodárenskou Věží 2, CZ-182 07 Praha 8

`zdenek@cs.cas.cz`

Odhady nových charakteristik polohy a disperze rozdělení, viz [1], lze zkonstruovat pomocí odhadnutých parametrů rozdělení. Jednou z možností jak tyto parametry odhadovat je modifikace zobecněné momentové metody, kterou je lze v řadě případů odhadovat přímo. Zobecněné momentové odhady jsou robustní pro rozdělení s dlouhými chvosty. Odhadnuté charakteristiky lze snadno porovnávat, což je výhodné v případě různým způsobem parametrizovaných rodin.

Reference

- [1] Fabián Z. (2008) New measures of central tendency and variability of continuous distributions, *Communication in Statistics, Theory Methods* **2**, 159–174.

Model dopravní špičky z mikroskopického pohledu

Lucie Fajfrová

ÚTIA AV ČR, Pod vodárenskou Věží 4, CZ-182 08 Praha 8

`fajfrova@utia.cas.cz`

Příspěvek bude věnován konzervativnímu částicovému systému, který zobecňuje známý (představený již Spitzerem [3]) částicový systém zvaný Zero-range proces. V tomto zobecnění povolíme přeskok více než jedné částice v jednom kroku. Soustředíme se na popis stacionárních stavů tohoto Markovského procesu a na konstrukci odpovídajícího couplingového procesu.

Zero-range proces byl použit jako model pro tok dopravy v [1]. Umožnění vícenásobných skoků pak vylepšuje právě interpretaci procesu jako modelu dopravní špičky, neboť vícenásobný přeskok zde odpovídá rozpadu kolony.

Ve speciálním případě, kdy intenzity přeskoků jsou konstantní (nezávislé na počtu přeskočivších částic ani na původním počtu částic), byl podobný proces - pojmenovaný "stick process" - studován v [2] v souvislosti s problémem nalezení délky nejdelší rostoucí podposloupnosti náhodné permutace a tzv. Hammersley-Aldous-Diaconis procesem.

Reference

- [1] Kaupuzs, J.; Mahnke, R.; Harris, R. J. (2005) Zero-range model of traffic flow. *Physical Review E* **72**, 056125.
- [2] Seppäläinen, T. (1996) A microscopic model for the Burgers equation and longest increasing subsequences. *Electronic J. Probab.* **1**, Paper 5, 1–51.
- [3] Spitzer, F. (1970) Interaction of Markov processes. *Adv. in Math.* **5**, 246–290.

Koziolův-Greenův model se zleva useknutými pozorováními

Michal Friesl

FAV ZČU, Univerzitní 22, CZ–306 14 Plzeň

friesl@kma.zcu.cz

Příspěvek pojednává o neparametrickém bayesovském odhadu funkce spolehlivosti v modelu s proporcionálním cenzorováním. Protože standardně se nepředpokládá, že by rozdělení cenzoru mohlo souviset s rozdělením doby života, musíme tvar neparametrických bayesovských odhadů pro tuto situaci upravit. Na minulém Robustu jsme se zabývali odhady z dat pouze cenzorovaných, tentokrát přidáme useknutí zleva.

Reference

- [1] Hyde J. *Testing survival under right censoring and left truncation*. *Biometrika* **64**, 225–230, 1977.
- [2] Koziol J. A., Green S. B. *A Cramér-von Mises statistic for randomly censored data*. *Biometrika* **63**, 465–474, 1976.
- [3] Kim Y. *Nonparametric Bayesian estimators for counting processes*, *Ann. Statist.* **27(2)**, 562–588, 1999.
- [4] Pawlitschko J. *Estimation in the Koziol-Green model with left truncated observations*, *Sankhya A* **62(1)**, 67–79, 2000.

Podmienený zákon veľkých čísel

Marian Grendár

KM FPV UMB, Tajovského 40, SK-974 01 Banská Bystrica

marian.grendar@savba.sk

Podmienený zákon veľkých čísel (Conditional Law of Large Numbers) je zaujímavé rozšírenie štandardného, učebnicového zákona veľkých čísel. Príspevok predstaví CLLN pre iid výber ako aj pre Pólyovu urnovú schému.

Odhady variančných parametrov v LRM s konečným diskretným spektrom

Martina Hančová

ÚMV PF UPJŠ, Jesenná 5, SK-040 01 Košice

martina.hancova@upjs.sk

Pri empirickej predikcii časových radov pomocou regresných modelov si prax žiada odhad variančných parametrov daného regresného modelu. V príspevku uvedieme nami navrhnuté dve varianty metódy prirodzeného odhadu — jeden založený na metóde najmenších štvorcov (OLSE) a druhý na najlepšej lineárnej nevychýlenej predikcii (BLUP) — pre odhadovanie variančných parametrov vo všeobecnej triede lineárnych regresných modelov (LRM) časových radov nazývaných LRM s konečným diskretným spektrom. Konečné pozorovanie tejto triedy modelov časových radov je špeciálnym prípadom lineárnych zmiešaných modelov. Doterajšie výsledky ukazujú, že odhady získané týmito metódami sú výhodné z výpočtového hľadiska (jednoduchý výpočtový algoritmus), ale aj z teoretického hľadiska (nekomplikované skúmanie teoretických vlastností odhadov) a existuje dôležité prepojenie medzi nimi a odhadmi získanými štandardnými metódami (napr. DOOLSE, MLE, REMLE).

Exponenciálny vyrovnávaní pro nepravidelné časové řady

Tomáš Hanzák

MFF UK, KPMS, Sokolovská 83, CZ-186 75 Praha 8

hanzak@karlin.mff.cuni.cz

Předpovědní a vyrovnávací metody typu *exponenciálního vyrovnávání* jsou s oblibou používány v praxi. V literatuře se v minulosti objevily také modifikace těchto metod pro časové řady pozorované v nepravidelných časových intervalech. V příspěvku jsou zdůrazněny problémy plynoucí v tomto případě z časové nepravidelnosti pozorování a naznačena možná řešení. Důsledně je rozlišována časová řada s chybějícími pozorováními a obecnější případ nepravidelné časové řady. Vedle již dříve publikovaných metod jsou prezentovány metody či jejich modifikace navržené autorem (exponenciální vyrovnávání řádu m , nepravidelně pozorovaný proces $ARIMA(0, 1, 1)$, modifikovaná Holtova metoda, Holt-Wintersova metoda modelující sezónnost pomocí goniometrických funkcí).

Algoritmy optimálneho navrhovania experimentov

Radoslav Harman

FMFI UK, KAMS, Mlynská dolina, SK – 842 48 Bratislava

harman@fmph.uniba.sk

Predstavme si, že sa chystáme vykonať experiment, pričom však nie sme len pasívni pozorovatelia, ale podmienky daného experimentu môžeme cielene ovplyvňovať; môžeme voliť pozíciu, alebo čas meraní, niektoré vlastnosti meraného objektu a podobne.

Zmyslom optimálneho navrhovania je stanoviť podmienky experimentu tak, aby sme z výsledkov vyťažili čo najviac informácie o neznámych charakteristikách nášho štatistického modelu. Ako však merať "množstvo informácie" získanej experimentom nie je jednoznačné; závisí od konkrétnych požiadaviek. Napríklad, cieľom môže byť minimalizovať disperziu odhadu konkrétneho parametra modelu, pričom hodnoty iných parametrov nás nezaujímajú (c -optimalita), alebo sa môžeme snažiť zorganizovať experiment tak, aby sme minimalizovali objem elipsoidu spoľahlivosti pre vektor všetkých parametrov (D -optimalita).

V prednáške sa zameriame na najjednoduchší model - lineárny regresný model s nekorelovanými pozorovaniami. Ukážeme si základné algoritmy, pomocou ktorých je možné optimálny spôsob experimentovania nájsť, pričom budeme kľásť dôraz na geometrickú interpretáciu týchto algoritmov a ich súvis so známymi optimalizačnými algoritmi. Napríklad, ukážeme si, že problém konštrukcie c -optimálneho návrhu experimentov je ekvivalentný problému nájsť prienik priamky s hranicou špeciálneho konvexného polyédra, alebo že algoritmus na výpočet D -optimálneho návrhu experimentu je možné použiť na konštrukciu elipsoidu minimálneho objemu obsahujúceho zadanú množinu bodov.

Model pro náhodné sjednocení kruhů se vzájemnými interakcemi

Kateřina Helisová

MFF UK, KPMS, Sokolovská 83, CZ – 186 75 Praha 8

helisova@karlin.mff.cuni.cz

Príspevek se zabývá modelem náhodné množiny dané konečným sjednocením kruhů, mezi nimiž se vyskytují vzájemné interakce a jejichž středy leží v omezené množině $S \subset \mathbf{R}^2$. Tento model je popsán hustotou pravděpodobnosti vzhledem k Booleovskému modelu (tj. modelu kruhů se vzájemně nezávislými středy a poloměry), která závisí na geometrických charakteristikách (např. plocha, obvod apod.) dané množiny. Prezentovány budou hlavně metody odhadu parametrů a testování vhodnosti modelu.

Funkcionální podmíněné kvantily rizikově neutrálních hustot

Zdeněk Hlávka

MFF UK, KPMS, Sokolovská 83, CZ-186 75 Praha 8

zdenek.hlavka@mff.cuni.cz

Kupní cenu $C_t(T)$ opce evropského typu v čase t s maturitou $T = t + \tau$, $\tau > 0$, lze zapsat jako střední hodnotu budoucího zisku $z(S_T)$ vynásobenou diskontním faktorem zohledňujícím bezrizikovou úrokovou míru r , tj. $C_t(T) = \exp\{-r\tau\}Ez(S_T) = \exp\{-r\tau\} \int_0^{+\infty} z(S_T)f_t(S_T)dS_T$, kde S_T je cena příslušného podkladového instrumentu, např. akcie, v čase T a tzv. *rizikově neutrální hustotu* $f_t(\cdot)$ lze interpretovat jako pravděpodobnostní hustotu náhodné veličiny S_T . Odhad rizikově neutrální hustoty bývá obvykle založen na pozorovaných cenách opcí.

V článku [2] je graficky znázorněna závislost změn tvaru odhadů rizikově neutrálních hustot na čase pomocí metody funkcionálních hlavních komponent [3]. Zde na tuto analýzu navážeme a pomocí neparametrické funkcionální analýzy [1] se pokusíme vyšetřit vztah mezi cenami akcií $S_{t+\nu}$, $\nu > 0$ a odhady rizikově neutrálních hustot $\hat{f}_t(\cdot)$. Kromě funkcionální neparametrické regrese ceny $S_{t+\nu}$ na funkce $\hat{f}_t(\cdot)$ použijeme i odpovídající podmíněné funkcionální kvantily.

Cílem příspěvku je zejména poskytnout základní přehled o metodách neparametrické funkcionální regrese a o problémech, se kterými se setkáme při jejich praktickém použití.

Reference

- [1] Ferraty F. a Vieu P. (2006) *Nonparametric Functional Data Analysis*. Springer, New York.
- [2] Hlávka Z. (2006) *Funkcionální hlavní komponenty pro odhady RNH*. Sborník prací 13. letní školy JČMF Robust 2006, 91–98.
- [3] Ramsay J.O. a Silverman B.W. (1997) *Functional Data Analysis*. Springer, New York.

Výprava do hlubin dat

Daniel Hlubinka, Lukáš Kotík, Ondřej Vencálek

MFF UK, KPMS, Sokolovská 83, CZ-186 75 Praha 8

daniel.hlubinka@mff.cuni.cz

Hlubka dat se stává uznávaným nástrojem pro statistickou analýzu mnohorozměrných dat. Hlavním účelem je vytvořit obdobu kvantilové funkce, která, vzhledem ke složitější geometrii vícerozměrných prostorů, nemá přirozenou obdobu ani pro dvourozměrné náhodné vektory. Svým přístupem se tato metoda, podobně jako kvantily pro náhodné veličiny, řadí do neparametrické statistiky. V poslední době se

objevují také pokusy o zobecnění pojmu hloubka na nekonečněrozměrné prostory, což je přístup použitelný například v analýze funkcionálních dat. Další slibně se rozvíjející oblastí je testování hypotéz založené na hloubce.

V současné době se všeobecně přijímá nahrazení lineárního uspořádání *větší/menší* opět lineárním (kvazi)uspořádáním *vnitřní/vnější* (kvazi- znamená, že více různých bodů může mít stejnou hodnotu). Hloubka bodu ve výběrovém prostoru je výsledkem *hloubkové funkce* spočítané pro dané rozdělení, případně pro empirické rozdělení založené na náhodném výběru. Body se stejnou hloubkou tvoří obrysy *vnitřních množin*, přičemž se obvykle hledá hloubka, pro kterou pravděpodobnost (podíl bodů ve výběru) vnitřní množiny má požadovanou hodnotu. Typicky je proto nutné počítat hloubku ve všech bodech, abychom mohli body uspořádat podle hloubky a rozhodnout o hraniční hodnotě hloubky.

Neexistuje jednoznačně přijímaná hloubková funkce, existují pouze obecná doporučení, jaké vlastnosti by hloubka měla mít. V minulosti bylo prozkoumáno mnoho typů hloubkových funkcí, jejichž vlastnosti se mohou dost lišit po teoretické i po výpočetní stránce. Podle přístupu je možné rozlišit několik základních skupin hloubek, v současnosti jsou nicméně nejpoužívanější *poloprostorová* hloubka, *simplexová* hloubka a na hodnotách hustoty založené *vrstevnice*.

V našem příspěvku ukážeme základní myšlenku hloubky dat, budeme diskutovat některé vlastnosti a zaměříme se na zobecněnou verzi poloprostorové hloubky, pro kterou základní vlastnosti, včetně silné konzistence, dokážeme. Dále se zmíníme o poměrně novém *směrovém* přístupu k hloubce. Ukážeme několik příkladů ilustrujících silné a slabé stránky hloubkových funkcí.

Estimating the scale parameter in Burr distribution

Erika Hönschová

Comenius University, Dept. of Applied Mathematics and Statistics, Mlynská dolina, SK-842 48 Bratislava 4

honschova@fmph.uniba.sk

An L -estimate of the scale parameter of the Burr distribution is presented. The parameter c of this distribution is assumed to fulfill the inequality $c > \frac{1}{2}$, which includes the Pareto distribution (the value $c = 1$) as a special case. The asymptotic normality of this estimate is proved and the order of convergence of the remainder term in the asymptotic linearity expansion of the presented statistic is also included.

Rekurentné vzorce pre počty niektorých druhov permutácií

Klára Hornišová

ÚM SAV, Dúbravská 9, 841 04 Bratislava

umerhornsavba.sk

Počty permutácií neobsahujúcich daný vzor, dvips (podpermutáciu) možno odvodiť z rekurentných vzorcov pre počty ich špeciálnejších podmnožín. Rozklady na špeciálnejšie podmnožiny sú výhodné najmä, ak je počet ich prvkov ohraničený, t.j. nerastie s dĺžkou permutácie. V prípadoch, keď sa takýto rozklad s rekurentným vzorcom dá nájsť, sú dôkazy jednoduchšie, ako pri dosiaľ najpoužívanejšej metóde vytvárajúcich stromov.

Testing the stability of the functional autoregressive process

Marie Hušková, Lajos Horváth, Piotr Kokoszka

MFF UK, KPMS, Sokolovská 83, CZ-186 75 Praha 8

marie.huskova@mff.cuni.cz

The talk concerns test procedures for detection of a change in a function autoregressive process. Particularly, we consider the functional time series $X_{i+1} = \Psi_i X_i + \varepsilon_i$, $i = 1, \dots, n$, where X_i are observations, ε_i i.i.d. errors and Ψ_i operators. We develop a test procedure for testing $H_0 : \Psi_1 = \dots = \Psi_n$ versus H_1 : there exists $m < n$ such that $\Psi_1 = \dots = \Psi_m \neq \Psi_{m+1} = \dots = \Psi_n$. Developed CUSUM type procedure uses functional principal component analysis. It is constructed to have a known asymptotic distribution, but asymptotic justification is very delicate. Theoretical properties are investigated. Finite sample performance is examined by an application to a data set.

Kvantily v sekvenční analýze bodu změny

Ondřej Chochola

MFF UK, KPMS, Sokolovská 83, CZ-186 75 Praha 8

chochola@karlin.mff.cuni.cz

Príspevek pojednáva o testovaní zmeny modelu pri postupne prichádzajúcich datech, kde máme k dispozícii tzv. tréningové obdobie, kedy jsou data beze zmeny. V modelu polohy je toto testovanie standardne zalozeno na L_2 , prípadne (pro výrazněji nesymetrická data lépe) na L_1 odhadoch. V příspěvku se zaměříme na kvantily parametru polohy, jakožto zobecnění mediánu. Teoretická část založená na asymptotice je doplněna simulacemi ukazující chování procedury v konečném čase. Dále se budeme snažit tento postup rozšířit na regresní model a uvažovat obecnější předpoklady.

Řešení problémů analýzy obrazu pomocí MAP a minima totální variace

Jiří Janáček

Ústav fyziologie AV ?R, Víde?ská 1083, CZ-142 00 Praha 4

janacek@biomed.cas.cz

MAP odhad konečného Gibbsovského náhodného pole dostaneme minimalizací aposteriorní energie. V analýze obrazu často používáme pole s apriorní energií danou totální variací. Model aditivního Gaussovského šumu vede na L^2 ztrátovou funkci, aposteriorní energie je ryze konvexní a příslušnou minimalizační úlohu lze řešit velmi efektivně iterační metodou založenou na nelineárních projekcích [1]. Poněkud robustnější L^1 ztrátová funkce dává energii pouze konvexní, kterou můžeme minimalizovat iterovaným hledáním optimální množiny obrazových elementů k pevně zvolené změně hodnoty [2]. Charakteristickou funkcí optimální množiny přitom dostaneme řešením příslušného binárního problému nalezením minimálního řezu ve vhodném neorientovaném grafu [3, 4]. Popsaná procedura poměrně dobře zachovává hrany a lze ji použít k potlačení obrazového šumu. Úlohy stereoskopie, stanovení optického toku a elastické registrace vyžadují odhad korespondence obrazových prvků, k jehož řešení můžeme také použít výše zmíněné metody, ovšem příslušná energie už nemusí být ani konvexní. Volba ztrátové funkce (L^2 vs. L^1) má výrazný vliv na vlastnosti řešení.

Reference

- [1] Chambolle A. An algorithm for total variation minimization and applications. *J. Math. Imaging Vis.* **20**, 89–97, 2004.
- [2] Murota K. On steepest descent algorithms for discrete convex functions. *SIAM J. Optim.* *14*, 699–707, 2003.
- [3] Greig D.M., Porteous B.T. a Seheult A.H. Exact maximum a posteriori estimation for binary images. *J. R. Statist. Soc.* **B51**, 271–279, 1989.
- [4] Kolmogorov V., Zabih R. What energy functions can be minimized via graph cuts? *IEEE TPAMI* **26**, 147–159, 2004.

O významu entropie

Martin Janžura

ÚTIA AV ČR, Pod vodárenskou Věží 4, CZ–182 08 Praha 8

janžura@utia.cas.cz

Označme $\mathcal{P}(\mathcal{X})$ množinu pravděpodobnostních měr na (pro jednoduchost) konečné množině \mathcal{X} . Pojem *entropie*, definovaný pro $P \in \mathcal{P}(\mathcal{X})$ výrazem

$$H(P) = \int -\log P \, dP = \sum_{x \in \mathcal{X}} -\log P(x) P(x),$$

a obecnější pojem *informační divergence* (*I -divergence*) pro dvojici $P, Q \in \mathcal{P}(\mathcal{X})$ daný jako

$$I(P|Q) = \int \log \frac{P}{Q} \, dP = \sum_{x \in \mathcal{X}} \log \frac{P(x)}{Q(x)} P(x)$$

pokud $P \ll Q$ (jinak $I(P|Q) = \infty$), jsou zcela přirozené charakteristiky pravděpodobnostních rozdělání, s velmi názornými a dobře interpretovatelnými vlastnosti.

Mimo jiné poskytují vhodný aparát pro snadný popis a práci s pojmy jako *nezávislost*, *markovost*, *postačitelnost*, *atd.*

Kromě oblastí, kde tyto charakteristiky původně vznikly a rozvíjely se, tj. *statistická fyzika a teorie informace*, ale mají i zásadní úlohu pro mnohé problémy teorie pravděpodobnosti a matematické statistiky, zejména pro:

- i) teorii statistického odhadování i testování hypotéz;
- ii) limitní věty teorie pravděpodobnosti;
- iii) popis mnohorozměrných rozdělení a náhodných procesů;
- iu) simulace.

Toto vše bude názornou formou prezentováno.

Testování normality ve vícerozměrném případě

Daniela Jarušková

ČVUT, stavební fakulta, kat. matematiky, Thákurova 7, CZ – 166 29 Praha 6

jarus@mat.fsv.cvut.cz

Příspěvek se bude zabývat zkušenostmi autorky s testováním normality ve vícerozměrném případě. Především se budeme zabývat testy založenými na „šikmosti a špičatosti“, a to jak Pearsonovým testem, tak i Neymanovým testem. Nakonec se budeme zabývat vícerozměrným Kolmogorovým Smirnovovým testem.

Reference

- [1] Fatalov V. R. Asymptotics of large deviation probabilities for Gaussian fields: Applications. *Izvetiya Natsionalnoi Akademii Nauk Armenii, Matematika* **28(5)**, 1993, 21 – 44.
- [2] Koziol J. A. An alternative formulation of Neyman's smooth goodness of fit tests under a composite alternatives. *Metrika* **34**, 1987, 17 – 24.
- [3] Thode H. C. jr. Testing Normality. Marcel Dekker, New York.
- [4] Tjurin Ju. N. Proverka gipotezy o normalnosti vyboroky bolshovo objema. *Teor. Veroy. i ee Primen.* **18(3)**, 1973, 651 – 655.

Využití MATLABu ve statistice a optimalizaci

Jaroslav Jirkovský

HUMUSOFT spol. s r.o., Pobřežní 20, CZ – 186 00 Praha 8

jirkovsky@humusoft.cz

MATLAB poskytuje svým uživatelům nejen grafické a výpočetní nástroje, ale i rozsáhlé knihovny funkcí spolu s výkonným programovacím jazykem čtvrté generace. Díky své architektuře je MATLAB určen zejména těm, kteří potřebují řešit početně náročné úlohy. Klíčovou vlastností, která patrně nejvíce přispěla k rozšíření MATLABu, je jeho otevřená architektura. MATLAB je úplný programovací jazyk,

takže uživatelé v něm mohou vytvářet funkce "šité na míru" pro jejich aplikace. Jednou z rozsáhlých aplikačních oblastí MATLABu je i statistika a optimalizace. Nástroji pro statistickou analýzu, modelování a vývoj algoritmů disponuje Statistics Toolbox. Obsahuje rozsáhlý soubor funkcí a interaktivní nástroje k analýze naměřených dat, modelování dat, simulaci systémů a vývoji statistických algoritmů. Základním nástrojem pro optimalizační úlohy je Optimization Toolbox. Přináší do MATLABu funkce a algoritmy pro standardní a rozsáhlé optimalizační úlohy. Tyto algoritmy řeší podmíněné a nepodmíněné spojité i diskrétní úlohy a zahrnují funkce pro lineární programování, kvadratické programování, nelineární optimalizaci, nelineární metodu nejmenších čtverců, řešení soustav nelineárních rovnic nebo víceúlohovou optimalizaci.

Vektorové autoregresní modely

Petr Jonáš

MFF UK, KPMS, Sokolovská 83, CZ – 186 75 Praha 8

`petr.jonas@centrum.cz`

Cílem této práce bylo shrnutí současného stavu problematiky vektorových autoregresních (VAR) modelů využívaných především v dnešní ekonometrii, popsat vhodný software pro příslušné aplikace těchto modelů a ilustrace jejich využití. V práci jsou dále popsány integrované procesy, princip kointegrace a VEC modely, které jsou vhodnou modifikací VAR pro kointegrované procesy. Věnuje se také problému Grangerovy a vícekrokové kauzality, analýze impulzních odezev a rozkladu chyby předpovědi v kontextu VAR modelů. Vše je doplněno ilustrativními příklady s reálnými daty zpracovanými v programu EViews.

Robustifikace hřebenové regrese - první pokusy

Tomáš Jurczyk

MFF UK, KPMS, Sokolovská 83, CZ – 186 75 Praha 8

`jurczyk@karlin.mff.cuni.cz`

Jeden z problémů, se kterými se můžeme při datové analýze setkat, je multikolinearita (téměř lineární závislost mezi regresory). Multikolinearita vede k nestabilitě řešení normální rovnice a může způsobit velký rozptyl odhadu regresních koeficientů. Jednou z metod, která se snaží tento problém řešit je hřebenová regrese. Tato metoda ale není imunní vůči přítomnosti kontaminace, která může multikolinearitu skrýt nebo ji naopak uměle vytvořit. Tento příspěvek se věnuje prvním jednoduchým pokusům o robustifikaci metody hřebenové regrese. V práci jsou představeny dva návrhy založené na robustní metodě nejmenších vážených čtverců (LWS).

Joseph-Émile Barbier a stereologie v XIX. století

Anna Kalousová

FEL ČVUT, kat. matematiky, Technická 2, CZ-166 27 Praha 6

kalous@math.feld.cvut.cz

Stereologie je statistická inference z geometrických výběrů, jejichž dimenze je obvykle nižší než dimenze zkoumaného objektu. První pravidla intuitivně zavedli M.A. Delesse v roce 1847 a A. Rosiwal v roce 1898. Počátky stereologie jsou většinou kladeny do druhé poloviny 20. století. Základní stereologické vztahy však najdeme již v článku J.-É. Barbiera (1839–1889) *Note sur le problème de l'aiguille et le jeu du joint couvert*, publikovaném v roce 1860.

J.-É. Barbier navštěvoval během studií na École normale supérieure, kde byl žákem Josepha Bertranda (1822–1900), také přednášky Gabriela Lamé (1795–1870) na Sorbonně. Tam se seznámil s Buffonovou úlohou o jehle a některými jejími zobecněními (jehla nahrazena kruhem, elipsou, pravidelným mnohoúhelníkem). V uvedeném článku nejprve shrnul všechny uvedené příklady do jedné obecné teorie a potom odvodil střední hodnotu počtu průsečíků libovolné rektifikovatelné křivky s čárovým systémem rovnoběžných ekvidistantních přímk. Uvědomil si, že vztah platí i v případě, kdy čáry nejsou rovnoběžné ani ekvidistantní, ba ani přímé. Stačí, aby v každém čtverci o jednotkovém obsahu byl součet délek těchto čar stejný (ekvivalentně pro střední hodnotu tohoto součtu). Dnes bychom řekli, že je známá délková intenzita čar. Analogicky pokračoval i v prostoru, odvodil vztahy pro výpočet střední hodnoty počtu průsečíků vlákna určité délky s plochou o dané plošné intenzitě, plochy určitého obsahu s vláknem o dané délkové intenzitě a střední hodnoty součtu délek průsečnic plochy určitého obsahu s plochou o dané plošné intenzitě. V těchto dvojicích interagujících geometrických objektů stačí znát intenzitu jednoho objektu (testovací systém) k odhadu intenzity objektu druhého (zkoumaný objekt). Jeho výsledky zůstaly nepovšimnuty a následně byly odvozeny až o řadu let později, viz např. S.A. Saltykov (1945), L.A. Santaló (1953), R. Miles (1972).

J.-É. Barbier trpěl duševní poruchou, kvůli které na patnáct let zmizel z vědeckého světa. Pak mu jeho bývalý učitel J. Bertrand, který byl v té době tajemníkem Akademie věd, zajistil nevelký příjem z nadace spojené s Akademií. Barbier napsal víc než deset dalších článků, žádný z nich už ale nebyl o geometrické pravděpodobnosti. Barbier je obecně znám jen svým příspěvkem k isoperimetrickému problému v rovině, konkrétně větou dokázanou ve výše uvedeném článku: *Všechny oblasti (obrazce) konstantní šířky (průměru) w mají obvod délky πw .*

Levy processes and balayage theory

Andrea Karlová

MFF UK, KPMS, Sokolovská 83, CZ-186 75 Praha 8 a

ÚTIA AV ČR, Pod vodárenskou Věží 4, CZ-182 08 Praha 8

karlova@karlin.mff.cuni.cz

The classical potential theory provides a very close connection between theory of probability and mathematical calculus. A very known is a Feymann-Kac theorem which very formally speaking allows us to interpret the solution of the partial differential equation of a specific form as an expected value of the functional of Brownian motion. The link between these two branches of mathematics is however much wider and there are four equivalent views on the potential theory via Hunt processes, sub-markov semigroups, families of harmonic kernels and balayage space (Theorem 8.1 in Bliedtner, Hensen [2]). In the talk we focus only on the particular examples of this general theory and shows its applications on the small subclass called Levy processes. The later covers large class of stochastic processes which however still owns very nice and handable properties.

Reference

- [1] Applebaum D. (2004). *Levy processes and stochastic calculus*. Cambridge University Press.
- [2] Bliedtner J., Hensen W. (1986). *Potential Theory*. Springer-Verlag, Berlin.

Intervalový odhad parametru p binomického rozdělení: Co je (relativně) nového?

Jan Klaschka

ÚI AV ČR, Pod Vodárenskou Věží 2, CZ – 182 07 Praha 8

klaschka@cs.cas.cz

Když jsem v r. 2006 mluvil o intervalových odhadech parametru p binomického rozdělení na Robustu poprvé, netušil jsem, že jen několik dní nato narazím na webu na pracovní verzi článku [1], který mi znovu potvrdí, že v souvislosti s daným tématem se stále lze dočkat překvapení, a přiměje mě doplnit první přednášku “druhým dílem”.

V práci [2] jsem se dotkl potíží a dilemat pramenících z toho, že pravděpodobnost pokrytí skutečné hodnoty parametru p binomického rozdělení (a obecně parametrů diskrétních rozdělení) konfidenčním intervalem se z principu nemůže na celém parametrickém prostoru shodovat s nominální hladinou spolehlivosti. Problém se sice dá vyřešit pomocí *znáhodněných konfidenčních intervalů*, ale spíše jen teoreticky: Uživatele statistiky sotva přimějeme brát vážně interval, jehož meze – pro jedna a tatáž empirická data – jsou dnes takové a zítra onaké podle toho, jaké číslo právě vydal (pseudo)náhodný generátor.

V současném příspěvku se budu věnovat jinému pokusu, jak dosáhnout shody s nominální hladinou spolehlivosti: Jedná se o tzv. *fuzzy konfidenční intervaly*, zavedené Geyerem a Meedenem [3]. (Agresti a Gottard na ně zajímavě navázali v již citovaném článku [1].)

Fuzzy konfidenční interval je funkce h (s hodnotami v $[0, 1]$), z níž se dá odečíst, jak by vypadaly znáhodněné konfidenční intervaly pro všechny možné hodnoty

náhodných vstupů. Samotná funkce h na náhodných vstupech nezávisí, což by mohlo fuzzy konfidenčním intervalům ve srovnání s intervaly znáhodněnými dávat větší šanci, že se uchytí v praxi.

Reference

- [1] Agresti A., Gottard A. (2007) *Nonconservative exact small-sample inference for discrete data*. Computational Statistics & Data Analysis **51**, 6447–6458.
- [2] Klaschka J. (2006) *O intervalových odhadech pravděpodobností, zvláště malých*. In: ROBUST 2006 (Antoch J., Dohnal G., eds.), JČMF, Praha, 119–126.
- [3] Geyer C.J., Meeden G.D. (2005) *Fuzzy and randomized confidence intervals and p-values*. Statistical Science **20**, 358–366.

Combined logistic regression estimator for the case-cohort proportional odds model

Petr Klášterecký, Michal Kulich

MFF UK, KPMS, Sokolovská 83, CZ–186 75 Praha 8

klaster@karlin.mff.cuni.cz

The Prentice's case-cohort design became very popular for designing large epidemiological studies and all usual survival regression models, including the proportional odds model, currently have their case-cohort variants. However, most of the estimators described in the literature utilise weighting individual contributions to the estimating function by inverse sampling probabilities, which leads to severe performance problems in studies where the outcome is a rare event.

In this contribution we identify common problems of the existing case-cohort estimators and present a new estimator of regression parameters in the proportional odds model showing better properties than traditional estimators. Our estimator is based on estimating regression parameters repeatedly by means of logistic regression models and combining these estimators into a new one. Due to the assumption of a proportional odds model the parameter values remain unchanged over time and all logistic regression estimators therefore estimate the same quantity. The whole process works as if a sequence of case-control studies was performed and therefore no weighting is necessary. We also discuss the choice of optimal times for performing logistic analyses, computing the optimal linear combination of the individual estimators and show asymptotic properties of the combined estimator.

Although this approach utilises certain features specific to logistic regression and proportional odds models, it also works well for data following the Cox proportional hazards model whenever the overall probability of an event is low.

R balíček pro odhad jedno i vícerozměrných rozdělení pomocí normálních směsí, případně s intervalově cenzorovanými daty

Arnošt Komárek

MFF UK, KPMS, Sokolovská 83, CZ – 186 75 Praha 8

arnost.komarek@mff.cuni.cz

V příspěvku bude představen nový R-kový balíček určený k odhadu jedno i vícerozměrných rozdělení pomocí normálních směsí s tím, že data mohou být intervalově cenzorována. Implementované metody jsou založeny na odhadu parametrů směsi pomocí Markov chain Monte Carlo (MCMC) simulace. Počet komponent směsi je buď jedním z odhadovaných parametrů modelu (Green, 1995, *Biometrika*) nebo je hledán minimalizací deviančního informačního kritéria (DIC, Spiegelhalter et al., 2002, *JRSS B*). Výklad teorie bude kombinován s ukázkami možností představovaného R-kového balíčku.

Směrové kvantily pro vícerozměrná data

Lukáš Kotík

ÚTIA AV ČR, Pod vodárenskou věží 4, CZ – 182 08 Praha 8

kotik@utia.cas.cz

Máme-li jednoznačně určený parametr polohy (střed) pro vícerozměrná data, můžeme přejít k hypersférickým souřadnicím s počátkem v tomto středu. Směrový kvantil v daném směru pak definujeme jako jednorozměrný kvantil rozdělení rádiusu podmíněného úhly, které určují směr ze středu. To můžeme provést pro každý směr a přechodem zpět do kartézských souřadnic získáme konfidenční region.

Pro určení středu zpravidla použijeme nějaké vícerozměrné rozšíření jednorozměrného mediánu (např. založeného na hloubce). Výběrový směrový kvantil můžeme odhadnout po přetransformování výběru do hypersférických souřadnic pomocí kvantilové regrese nebo jádrového vyhlazování. Příspěvek je zaměřen na základní vlastnosti konfidenčních množin získaných pomocí směrových kvantilů a na možnosti jejich odhadu.

Text mining of articles in metabolomics research

Milena Kovářová

Institute of Physical Biology, University of South Bohemia, Zámek 136, CZ – 373 33 Nové Hradky

kovarova@greentech.cz

More than 80% of information is stored in textual form. A large number of papers appear every day. It is beyond power of anybody to go through even a part of the published works. A use of search machines yields excessive numbers of results and

does not make a problem simpler. There are additional difficulties in biochemical research such as an abundance of terms with ambiguous nomenclature, acronyms and abbreviations. That is the reason for a boom of text mining methods in this area. Text mining methods transfer unstructured textual data into a structured data matrix, which can be evaluated with data mining methods.

Základy štatistiky s fuzzy dátami

Pavol Král', Lenka Lašová, Marek Ďurica

EF UMB, Tajovského 10, SK – 975 90 Banská Bystrica;

FPV UMB, Tajovského 40, SK – 974 01 Banská Bystrica

pavol.kral@umb.sk, lasova@fpv.umb.sk, durica@fpv.umb.sk

Pod fuzzy dátami rozumieme náhodné vágne dáta vyjadrené prirodzeným jazykom. Cieľom príspevku je zhrnutie súčasného stavu štatistiky s fuzzy dátami z pohľadu teórie fuzzy množín, pričom sa zameriavame najmä na popisnú štatistiku a základy inferenčnej štatistiky.

Proč nám nerozumějí?

Milena Kvaszová

MÚ AV ČR, Žitná 25, CZ – 115 67 Praha 1

milena.sp@centrum.cz

Príspevek je venovaný niektorým príčinám obtíží pri vyučovaní teórie pravdepodobnosti a matematickej štatistiky. Výchozím bodom je model poznávacího procesu [1], ktorý vytvoril Jean Piaget (1896–1980). Podľa tohoto modelu je možné vo vývoji a chápaní určitej matematickej teórie rozlíšiť tri štádia, ktoré Piaget pojmenoval *INTRA*, *INTER* a *TRANS*. Ve štádiu *INTRA* študent dokáže sledovať jednotlivé jevy, ale není ještě schopen chápat a rozpoznat jejich souvislosti. Ve štádiu *INTER* študent správne chápe souvislosti mezi konkrétními jevy, ale ještě je nedokáže transformovat a vidět z různých pohledů. Ve štádiu *TRANS* študent již plně porozumí analogiím, transformacím atd., viz též [2], [3].

Problémy, ktoré bránia porozumeniu pravdepodobnosti a štatistiky u žiakov, často pramenia z toho, že pojmy a poznatky z úrovne *INTER* a *TRANS* sa študentom prezentujú bez toho, aby sa jim umožnilo prejsť prvým štádiom, tj. *INTRA*. Odtiaľ pak vyplývá požiadavka na zařazení elementární výuky o nejistých a omezeně kauzálních jevech ji v základním školství. Podobně je nezbytné včas naučit žáky samostatně zpracovávat jednoduchá data, a to především ta, která se jich bezprostředně týkají.

Reference

- [1] Piaget J., Garcia R. (1989) *Psychogenesis and the History of Science*. Columbia University Press, New York.
- [2] Kvasz L. Why don't they understand us? *Science and Education* **6**, 263–272.

- [3] Kvasz L. (2008) *Patterns of Change, Linguistic Innovations in the Development of Classical Mathematics*. Birkhäuser Verlag AG, Basel.

Testy hypotéz o parametrech reškálovaného Wienerovho procesu s posunutím

Andrea Kvitkovičová

MFF UK, KPMS, Sokolovská 83, CZ – 186 75 Praha 8

akvitkovicova@centrum.cz

V příspěvku sa zaoberáme testovaním hypotéz o parametrech a a b náhodného procesu $\{bw_t + at; t \geq 0\}$, kde $\{w_t; t \geq 0\}$ je Wienerov proces, $a \geq 0$ a $b > 0$. Testy sú založené na čase, kedy proces prvýkrát dosiahne vopred zvolenú kladnú hranicu. Dôraz kladieme na testy hypotéz o parametri posunutia a , predovšetkým o jeho nulovosti. Skonstruujeme rovnomerne najsilnejšie nestranné testy jedného parametra v prítomnosti parametra rušivého. Ďalej skonstruujeme rovnomerne najsilnejšie testy hypotéz o parametri posunutia a za predpokladu, že rozptyl b^2 je známy. Touto situáciou sa potom zaoberáme podrobnejšie. Zostrojíme tiež testy obidvoch parametrov súčasne založené na Rényiho divergenciách.

Model pro předávání genů

Petr Lachout

MFF UK, KPMS, Sokolovská 83, CZ – 186 75 Praha 8;

ÚTIA AV ČR, Pod vodárenskou věží 4, CZ – 182 08 Praha 8

petr.lachout@mff.cuni.cz

Jednou z otázek řešených v současnosti v biologii je optimální výběr ze zadané populace. Vstupem je rodokmen sledované populace a znalost genů v rodičovské generaci. Geny celé populace nejsou známy. Je však znám mechanismus předávání genů z rodičů na děti. Tento mechanismus je náhodný a tak umožňuje určit pravděpodobnostní rozdělení genů u každého jedince. Pro dva jedince spočteme jejich genovou vzdálenost zvanou též „diversifity“ jako pravděpodobnost, že jejich potomek bude mít oba geny shodné. Dále sledujeme zisk (gain), což je ukazatel, který dokážeme změřit u každého jedince v uvažované populaci.

Pro vybranou podpopulaci vypočteme její rozrůzněnost, jako průměr genových vzdáleností přes všechny dvojice jejích jedinců, a její zisk jako průměr zisků všech jejích jedinců. Úkolem je vybrat podpopulaci s maximální rozrůzněností a s maximálním ziskem. Tento příspěvek představí model vhodný pro tuto situaci.

Analýza rozmístění elektronických podatelů obcí v ČR

Radka Lechnerová^[1], Tomáš Lechner^[2]

^[1] SVSEŠ, s.r.o., Lindnerova 575/1, CZ – 180 00 Praha 8

^[2] VŠE, NF, katedra práva, nám. W. Churchilla 4, CZ – 130 67 Praha 3

radka.lech@seznam.cz, lechner@triada.cz

E-Government jako jeden z prostředků transformačního procesu veřejné správy vytváří nové elektronické veřejné služby. Rozsah poskytování těchto služeb však dosud není na očekávané úrovni a neodpovídá plně platné legislativě, což vyvolává celou řadu otázek, proč tomu tak je? V článku zkoumáme aktuální fungění elektronické podatelny úřadů územních samosprávných celků, a to zejména s ohledem na jejich prostorové rozmístění v rámci České republiky, které analyzujeme prostřednictvím sumárních statistik bodových procesů. Tímto způsobem hledáme odpověď na otázku, zda mezi obcemi, které provozují elektronickou podatelnu, existují nějaké interakce vyplývající z jejich vzájemných poloh nebo zda jsou elektronické podatelny rozmístěny v náhodně vybraných obcích. Získané výsledky jsou podstatné pro vytvoření celkového obrazu průběhu procesu E-Governmentu v České republice.

Detekce defektů ve tkanině pomocí vícerozměrných regulačních diagramů

Aleš Linka, Maroš Tunák, Vladimír Bajzík

KHT, FT TUL, Studentská 2, CZ-461 17 Liberec 1

ales.linka@tul.cz, maros.tunak@tul.cz, vladimir.bajzik@tul.cz

Příspěvek se věnuje hodnocení tkaných textilií z pohledu detekce běžných defektů a náhodných odchylek od pravidelné struktury, tj. charakteristik ovlivňujících kvalitu tkanin. V příspěvku je studován statistický přístup, který je založen na statistických charakteristikách druhého řádu získaných z matice plošných šedotónových závislostí. Pomocí metody CART byly stanoveny texturní charakteristiky s největší diskriminační silou, které pak byly využity pro detekci defektů v texturních obrazech tkanin. K automatické detekci defektních oblastí byla použita technika současného monitorování více texturních charakteristik pomocí vícerozměrných regulačních diagramů. Navrhované metody jsou vhodné pro detekci běžných typů defektů, což potvrzují výsledky získané jak na simulovaných strukturách tak na obrazech reálných textilií.

Spline models with change-points applied to seroprevalence data

Matúš Maciak

MFF UK, KPMS, Sokolovská 83, CZ-186 75 Praha 8

mmatthew@matfyz.cz

In this work we focus on a problem of estimating an unknown regression function based on spline approach, while taking into account also some sudden changes – change-points.

Firstly, we have discussed some basic methods and their properties for a simple spline approach and afterwards we have tried to implement an algorithm which

would allow for change-points occurrences at some certain, pre-defined points – so called knot points.

As far as the mesh of knots is given in advance before we actually fit the unknown regression curve there was a need to come up with strategy how to define knot points where one can expect change-points. We have discussed two different strategies which are plausible here: an intuitive knots placement based on some exploratory analysis and also jump-detection algorithm for a local polynomial modelling.

The proposed methods are applied to seroprevalence data of B19 parvovirus in Belgium population. Different models have been fitted and compared with respect to several statistical criteria. At the end we also discuss some advance modifications, like robust approach.

Ratio type statistics for detection of changes in linear regression models

Barbora Madurkayová

MFF UK, KPMS, Sokolovská 83, CZ–186 75 Praha 8

madurka@karlin.mff.cuni.cz

Linear regression models are relatively frequently used models in statistical analysis, with possible applications e.g. in econometrics, biology or climatology. We study procedures for detection of changes in parameters of linear regression models, particularly test procedures based on ratio type statistics.

To the most remarkable advantages of ratio type test statistics belongs the fact, that when computing the test statistics, it is not necessary to estimate variance of the underlying model. This property makes the ratio type statistics a suitable alternative of classical (non-ratio) statistic – most of all in situations, when it is difficult to find a suitable variance estimate.

We focus on the the properties of ratio type test statistics for detection of changes in linear regression models and demonstrate these properties on simulated data.

Příběh z cyklu „Statistický pohled na měření v české a světové literatuře“

Jaroslav Marek

KMAAM Přf UP, Tomkova 40, CZ–779 00, Olomouc – Hejčín

marek@inf.upol.cz

V příspěvku se pokusíme s využitím statistiky posoudit ústřední téma knihy Tajný ostrov templářů. Autoři E. Haagensen a H. Lincoln uvádí, že středověké kostely na ostrově Bornholm vytváří posvátnou konfiguraci. Ze schematu v knize lze sestavit regresní model s podmínkou typu II a ze známých souřadnic kostelů lze odhadnout střed konfigurace. Nalezený odhad pak může posloužit pro posouzení věrohodnosti tvrzení autorů.

Reference

- [1] Kubáček L., Kubáčková L. (2000) Statistika a metrologie. Vydavatelství UP, Olomouc.
 [2] Haagenzen E., Lincoln H. (2003) Tajný ostrov templářů, BB Art Praha.

O odhadu vzájemné informace

Tomáš Marek, Petr Tichavský, Georges Darbellay, Jaroslav Franěk

ÚTIA AV ČR, Pod vodárenskou Věží 4, CZ-182 08 Praha 8

marek@utia.cas.cz

Vzájemná informace je rozšířenou mírou vzájemné závislosti jednotlivých složek vícerozměrných náhodných vektorů. Časté uplatnění nachází především v inženýrských aplikacích. Pro dvojrozměrný náhodný vektor $(X, Y)^T$ se sdruženou hustotou $f_{X,Y}$ a marginálními hustotami f_X a f_Y je vzájemná informace $I(X, Y)$ náhodných veličin X a Y definována jako

$$I(X, Y) = \int_{\mathbf{R}^2} f_{X,Y}(x, y) \log \frac{f_{X,Y}(x, y)}{f_X(x) f_Y(y)} dx dy.$$

Metody odhadu vzájemné informace většinou vychází ze známého vztahu mezi vzájemnou informací a entropiemi příslušných rozdělení

$$\begin{aligned} I(X, Y) &= H(X) + H(Y) - H(X, Y), \\ &= H(X) - H(X|Y), \\ &= H(Y) - H(Y|X). \end{aligned}$$

Hodnotu $I(X, Y)$ je možné odhadnout také přímo. Na příkladech různých typů dvojrozměrných rozdělení srovnáme některé dostupné metody odhadu vzájemné informace.

Multiple changes in coefficients of autoregressive models

Miriam Marušíaková

MFF UK, KPMS, Sokolovská 83, CZ-186 75 Praha 8

maruskay@gmail.com

We consider an autoregressive model of order p with m structural changes

$$y_i = \mathbf{x}_i' \boldsymbol{\beta}_j + e_i \quad t_{j-1} < i \leq t_j, \quad j = 1, \dots, m+1$$

where t_1, \dots, t_m are the change points, $\mathbf{x}_i = (1, y_{i-1}, \dots, y_{i-p})'$ are lagged observations, $\boldsymbol{\beta}_j = (\beta_{j0}, \beta_{j1}, \dots, \beta_{jp})'$ are unknown regression parameters in the j th segment, $j = 1, \dots, m+1$, and e_1, \dots, e_n are independent identically distributed errors.

The change points t_1, \dots, t_m and their number m are in practice mostly unknown. We deal with F type tests for detection of changes, namely tests for no change against k changes or maximum k changes. The tests were proposed by [2] for linear regression models with non-trending regressors. Using results of [4], we show the tests can be applied also to autoregressive models. Approximations to the corresponding critical values can be obtained through the limiting distribution of the test statistic under the null hypothesis. The other possibility which we propose is to apply residual bootstrap with or without replacement, see, e.g., [1] and [3]. We prove the resampling methods provide asymptotically correct approximations to the critical values of the given tests. We conduct a simulation study to see how the test procedures work in finite samples.

Reference

- [1] Antoch J. and Hušková M. (2001). Permutation tests in change point analysis. *Statist. Probab. Letters* **53**, 37–46.
- [2] Bai J. and Perron P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica* **66**(1), 47–78.
- [3] Hušková M., Kirch C., Prášková Z. and Steinebach J. (2008). On the detection of changes in autoregressive time series II. Resampling procedures. *J. Statist. Plann. Inference* **138**, 1697–1721.
- [4] Hušková M., Prášková Z. and Steinebach J. (2007). On the detection of changes in autoregressive time series I. Asymptotics. *J. Statist. Plann. Inference* **137**, 1243–1259.

Robustní statistické návrhy

Jiří Michálek

ÚTIA AV ČR, Pod vodárenskou Věží 4, CZ–182 08 Praha 8

michalek@utia.cas.cz

Statistické návrhy neboli DOE (Design of Experiments) je oblast aplikované matematické statistiky, které bohužel u nás není věnována patřičná pozornost jak z hlediska teorie, tak i z hlediska uplatnění v praxi, kde často se jedná o jediný možný efektivní způsob, jak porozumět vlivu fixních či náhodných faktorů na sledovanou veličinu, kterou často v praxi bývá nějaký jakostní znak, jehož chování je nutné udržet pod kontrolou. Velice častým problémem, který se dá řešit pomocí DOE je nalezení vhodného nastavení verzí fixních faktorů takovým způsobem, aby sledovaná veličina, tzv. response, splňovala nějaké požadavky, např. požadovaná střední hodnota či úroveň variability. Velice často v problémech tohoto typu hrají náhodné faktory roli šumu, který nelze odstranit, je možno pouze tyto faktory měřit, nikoliv ovlivňovat či nastavovat jejich verze na rozdíl od fixních (nenáhodných) faktorů. Vzniká tedy problém, jak dosáhnout požadovaného chování sledované výstupní veličiny, které by bylo robustní právě vůči vlivu šumu či jiných rušivých faktorů. Do tohoto schématu se vejde většina problémů s nastavením výrobního procesu, kde

produkt, z něj vycházející, má splňovat předem dané požadavky, např. na výskyt neshodných výrobků. Na začátku osmdesátých let přišel japonský odborník Taguchi s konstrukcí statistických návrhů, do kterých zabudoval i postup na zvládnutí rušivých faktorů. Rozdělil faktory na říditelné, jejichž verze lze nastavovat, a rušivé. Celkový experiment pak obsahuje počet běhů, který se získá krížením dvou tzv. ortogonálních oblastí. Jedna oblast obsahuje běhy pro říditelné faktory, druhá oblast obsahuje uvažované stavy rušivých faktorů. Pro vyhodnocení experimentu Taguchi používá kritéria založená na poměru signál/šum. Nejčastěji používaná kritéria mají tvar „větší je lepší“, či „menší je lepší“ nebo „nominální je nejlepší“. Vyhodnocení se nejčastěji provádí pomocí analýzy rozptylu. Taguchiho přístup je často kritizován především ze strany matematických statistiků, např. za ignorování možných interakcí mezi říditelnými a rušivými faktory nebo též za používání kritérií založených na poměru signál/šum.

Příspěvek se zabývá modifikací Taguchiho přístupu, která vychází z častých požadavků na chování výrobního procesu. Pro jednoduchost si představme, že na produktu je sledován jakostní znak spojitého charakteru, např. rozměr, na který jsou kladeny požadavky ve formě např. předepsaných ukazatelů způsobilosti C_p a C_{pk} . Jinými slovy to znamená, že sledovaný jakostní znak lze popsat normálním rozdělením s předepsanou střední hodnotou, např. prostředek tolerančního rozmezí, a s předepsanou směrodatnou odchylkou. Objevuje se tedy problém, jak nastavit verze říditelných faktorů, aby se chování jakostního znaku „co nejlépe“ drželo požadovanému normálnímu rozdělení za přítomnosti rušivých faktorů. Nabízí se tedy myšlenka měřit tuto „vzdálenost“ měřit pomocí některé z bohaté třídy f -divergencí. F -divergence co by měly příbuznosti mezi pravděpodobnostními rozděleními jsou ve matematické statistice velice úspěšně používány a jejich teorie je hluboce propracována.

Myšlenka je tedy namísto Taguchiho kritéria „nominální je nejlepší“ použít některou vhodnou f -divergenci, např. mezi dvěma normálními rozděleními, kde jedno představuje požadavek na chování jakostního znaku, a druhé vychází z naměřených dat získaných během provádění experimentu. V příspěvku bude uveden též příklad robustního návrhu, který bude analyzován pomocí vhodně zvolené modifikace kritéria založené na f -divergenci.

Change point detection by basis pursuit

Jiří Neubauer^[1], Vítězslav Veselý^[2]

^[1]KE FEM UO, Kounicova 65, CZ–612 00 Brno

^[2]KAMI ESF MU, Lipová 41a, CZ–602 00 Brno

neubauer@email.cz, vesely@econ.muni.cz

The contribution deals with use of overcomplete models and sparse parameter estimation for change point detection in one-dimensional stochastic processes. These processes are estimated by heavyside functions. The BASIS PURSUIT algorithm is used to get sparse parameter estimation. The mentioned method of change point

detection in stochastic processes is compared with standard methods by simulations.

Porovnání metod odhadu vyhlazovacího parametru při jádrových odhadech hustoty

Jan Orava

PrF MU, OAM, Kotlářská 2, CZ-611-37 Brno

orava@mail.muni.cz

Jádrový odhad pravděpodobnostní hustoty patří do skupiny neparametrických odhadů, tedy nepotřebujeme znát žádné apriorní informace o neznámé hustotě. Samotná konstrukce jádrového odhadu je velmi jednoduchá, ovšem kvalita výsledného odhadu závisí na několika významných faktorech. Právě vyhlazovací parametr má na kvalitu výsledného odhadu největší vliv a vhodná volba vyhlazovacího parametru je největším problémem při jádrovém odhadu hustoty. Příspěvek bude věnován srovnání účinnosti šesti různých metod odhadu vyhlazovacího parametru na simulovaných datech ze čtyř různých rozdělení pravděpodobnosti.

Detecting atoms in deconvolution

Jaroslav Pazdera

MFF UK, KPMS, Sokolovská 83, CZ-186 75 Praha 8

pazdera@matfyz.cz

In the ordinary deconvolution problem one wants to estimate the distribution of Y from the given sample of i.i.d. random variables X_1, \dots, X_n , where $X_i = Y_i + Z_i$, and where Z_i is supposed to be an error term with a known distribution. In the ordinary deconvolution problem it is usually assumed that Y has a continuous density. In the atomic deconvolution we suppose that the distribution of Y has an atom at a_0 .

We will propose an estimator for the location of the atom a_0 . And we will derive the asymptotic behavior of such estimator. We will use its relation to the asymptotic distribution of derivatives of an ordinary deconvolution kernel density estimator. This all will be done for a class of kernels with special properties. In the last part we will also present possible ways for the future development.

Ako lepšie porozumieť nelineárnej metóde najmenších štvorcov pomocou geometrie

Andrej Pázman

FMFI UK, KAMS, Mlynská dolina, SK-842 48 Bratislava

pazman@fmph.uniba.sk

Vo všeobecnosti budeme uvažovať nelineárny regresný model

$$\begin{aligned} y_i &= \eta(x_i, \theta) + \varepsilon_i; \quad \theta \in \Theta \subset \mathbb{R}^m \\ E(\varepsilon_i) &= 0, \quad \text{Var}(\varepsilon_i) = \sigma^2; \quad i = 1, \dots, N \end{aligned}$$

alebo vo vektorovom zápise

$$\begin{aligned} y &= \eta(\theta) + \varepsilon \\ E(\varepsilon) &= 0, \quad \text{Var}(\varepsilon) = \sigma^2 I \end{aligned}$$

kde $y = (y_1, \dots, y_N)^T$ je vektor dát, $\eta(\theta) = (\eta(x_1, \theta), \dots, \eta(x_N, \theta))^T$ a kde $\varepsilon = (\varepsilon_1, \dots, \varepsilon_N)^T$ je vektor chýb. Neznáme parametre $\theta = (\theta_1, \dots, \theta_m)^T$ sa odhadujú (nelineárnou) metódou najmenších štvorcov (MNS)

$$\hat{\theta} = \arg \min_{\theta \in \Theta} \sum_{i=1}^N [y_i - \eta(x_i, \theta)]^2$$

ktorá vďaka nelinearite modelu môže mať zaujímavé vlastnosti.

Obsah prednášky: Ukážka zložitosti zápisu istého nelineárneho modelu z aplikácií. Plocha stredných hodnôt a reparametrizácia modelu. Dotykový priestor a Gaussov-Newtonov algoritmus výpočtu MNS. Krivosť plochy stredných hodnôt a zlyhávanie MNS. Oblasť spoľahlivosti pre θ v modeloch s malou vnútornou krivosťou. Geometrické ohraničenia možnosti lineárizácie nelineárneho regresného modelu. Ako možno geometricky vysvetliť asymptotickú normalitu MNS odhadov. Geometrická klasifikácia nelineárnych modelov. Geometrické odvodenie „neasymptotickej“ hustoty odhadu $\hat{\theta}$. Problém neidentifikovateľnosti a „overlapping“.

Change detection in autoregressive time series

Jakub Pečánka

MFF UK, KPMS, Sokolovská 83, CZ-186 75 Praha 8;

Dept. of Stochastics, Vrije Universiteit Amsterdam

pecanka@gmail.com

Autoregressive time series models of a given order p have $p + 2$ parameters. The mean, the variance and the p autoregressive parameters. Detection of change in the value of any of these parameters is an important problem. This poster presents results from the theory of change point detection in autoregressive processes. The attention is given to statistics and tests based on the sums of weighted residuals, and the maximum likelihood type statistics. Main part of the poster presents computer simulations of the behavior of the studied statistics.

Zkušenosti se statistickým modelováním spotřeby zemního plynu

Emil Pelikán, Marek Brabec, Marek Malý, Ondřej Konár

ÚI AV ČR, Pod Vodárenskou Věží 2, CZ – 182 07 Praha 8

mbrabec@szu.cz, konar@cs.cas.cz, marek.maly@szu.cz, pelikan@cs.cas.cz

Úlohou statistického modelování spotřeby zemního plynu se kolektiv pracovníků ÚI AV ČR začal zabývat od roku 2003, kdy byl ve spolupráci se Západočeskou plynárenskou, a.s. vyvinut model pro odhad tzv. nevyfakturované dodávky zemního plynu. Práce na tomto tématu se intenzivněji rozeběhly v souvislosti s udělením grantové podpory GA AV ČR v rámci programu „Informační společnost“ na léta 2005–2009 č. 1ET400300513 a dále s účastí na projektu „Tvorba typových diagramů dodávky zemního plynu“, který je řešen ve spolupráci s partnery v České republice i na Slovensku. V příspěvku jsou shrnuty zkušenosti řešitelů, získaných zejména při překonávání překážek, které se vyskytly v průběhu návrhu, implementace a vlastního testování statistických modelů spotřeby zemního plynu. Příspěvek bude rozdělen do čtyřech hlavních částí, popisujících obecný úvod do problematiky, stav řešení ve světě, příklady použitých statistických modelů a faktory ovlivňující kvalitu modelu. Závěr příspěvku pak bude věnován vyhodnocení významu statistického modelování spotřeby zemního plynu a jeho dalším perspektivám.

Regression in Sobolev spaces using total least squares

Michal Pešta

MFF UK, KPMS, Sokolovská 83, CZ – 186 75 Praha 8

pesta@karlin.mff.cuni.cz

We propose a class of nonparametric estimators for the regression models based on total least squares over the sets of sufficiently smooth functions. The estimation takes place over the balls of functions which are elements of a suitable Sobolev space—special type of Hilbert spaces that facilitate calculation of the (total) least squares projection. The Hilbertness is allowing us to take projections and hence to decompose spaces into mutually orthogonal complements. Then we transform the problem of searching for the best fitting function in an infinite dimensional space into a finite dimensional optimization problem. The regression set-up proposed by Yatchew and Bos (1997) will be extended and combined with the total least squares approach introduced by Golub and van Loan (1980).

Reference

- [1] Golub G.H. and van Loan C.F. (1980) An analysis of the total least squares problem. *SIAM Journal on Numerical Analysis* **17**(6), 883–893.
- [2] Yatchew A. and Bos L. *Nonparametric least squares estimation and testing of economic models*. *Journal of Quantitative Economics* **13**, 81–131, 1997.

Poznámky k metodě POT

Jan Pícek

FP TUL, KAP, Studentská 2, CZ-461 17 Liberec

jan.picek@tul.cz

Jako vhodný model pro aplikace v analýze extrémních hodnot se používá zobecněné rozdělení extrémních hodnot, které odpovídá ve většině případů limitnímu rozdělení vhodně normovaného maxima sledované veličiny. Pokud pozorujeme veličiny v čase (např. teplota, srážky), potom se v této souvislosti uvažuje nejčastěji řada maximálních hodnot v jednotlivých rocích. Protože sledované období bývá obvykle krátké, je k dispozici pro analýzu poměrně málo hodnot.

Jednou z užívaných alternativ je proto metoda Peaks Over Threshold (POT), která uvažuje všechny hodnoty větší než zvolený threshold a vede k modelům založeným na zobecněném Paretově rozdělení. Z hlediska aplikací je důležitým a ne snadným problémem nastavení vhodného thresholdu. Cílem tohoto příspěvku je několik poznámek k jeho volbě na základě zkušeností, které získal autor s aplikacemi klimatologických dat.

Nejmenší vážené čtverce v příkladech

Pavel Plát

UK FSV, Smetanovo nábřeží 6, CZ-110 Praha 1

Plat@kmlinux.fjfi.cvut.cz

V příspěvku je na příkladech podrobně analyzováno chování metody nejmenších vážených čtverců (LWS_ψ) a výsledky porovnány s odhady regresních koeficientů vypočtených klasickou metodou nejmenších čtverců (LS), metodou nejmenších usekaných čtverců (LTS_h) a s metodou minimalizace h -té pořádkové statistiky kvadrátů reziduí (LMS_h – the least median of square). Výsledky demonstrují řadu žádoucích vlastností nejmenších vážených čtverců. LWS_ψ je odhad s vysokou odolností vůči kontaminaci dat „hrubými“ chybami. S kontaminací se dokáže vypořádat stejně dobře jako nejmenší usekané čtverce. Jeho vlastnosti můžeme navíc ovlivnit volbou váhové funkce. Díky možnosti volby spojitě váhové funkce pak mají nejmenší vážené čtverce menší podsouborovou citlivost než nejmenší usekané čtverce. Nabízí se zde srovnání s M -odhadem se spojitou ψ -funkcí. Oproti nim mají však nejmenší vážené čtverce tu výhodu, že jsou regresně a škálově ekvivariantní, což M -odhady obecně nejsou.

Moment estimates for spatial cluster point processes

Michaela Prokešová

MFF UK, KPMS, Sokolovská 83, CZ-186 75 Praha 8

michaela.prokesova@mff.cuni.cz

When estimating the parameters of spatial point process models we have two options. We can use the maximum likelihood estimation, which is asymptotically optimal but computationally very intense, because to obtain the numerical approximation of the likelihood (a closed form is typically not available) we need a large number of simulations from the estimated models. Especially for inhomogeneous models this procedure is practically unusable. The other option is to use moment methods, which are not optimal but are computationally easier and often precise enough. Recently several new moment methods for the estimation of homogeneous and inhomogeneous cluster processes were introduced. In the contribution we give an overview of the classical (minimum contrast) and new (composite likelihood, estimating equations, Palm likelihood) estimation methods, including a simulation study which compares the behaviour of the discussed methods for particular models.

Fuzzy míry neurčitosti měření za nejistoty a v podmínkách vágnosti

Zdeněk Půlpán, Michal Čihák, Jaroslava Bomberová

PdF UHK, Rokitanského 62, CZ-500 03 Hradec Králové 3

zdenek.pulpan@uhk.cz, michal.cihak@uhk.cz, jaroslava.bomberova@uhk.cz

Měření za nejistoty a v podmínkách vágnosti má svá omezení, která jsou dána krom jiného i absencí prototypu jednotky měřené vlastnosti. Měření pak nemá tu univerzálnost, kterou má např. měření délky nebo hmotnosti. V případě absence „jednotkové vlastnosti“ je nutné zavést „prototypy“ vlastností, se kterými se pak porovnávají vlastnosti měřeného vzorku. Příspěvek naznačuje možnosti odhadu neurčitosti a informace takových „měření“.

Moderní přístupy k testování periodicity v časových řadách

Veronika Ročková

MFF UK, KPMS, Sokolovská 83, CZ-186 75 Praha 8

rockova.ver@centrum.cz

V příspěvku je diskutován problém testování periodicity v časových řadách. Hlavní pozornost je věnována přístupu, který využívá informační kritéria na odhad řádu regresního modelu. Jsou prezentovány výsledky některých simulací, které naznačují, že testy pomocí kritérií dosahují lepších výsledků než klasický Fisherův test. Rozdíl je patrný především u modelů, ve kterých je periodicitu překryta náhodnou složkou. Tento přístup umožňuje nejen efektivně testovat přítomnost periodicity, ale také odhadnout, kolik periodických tendencí řada má, a tím získat komplexní obraz o periodickém chování řady.

Probabilistic modelling of the sleep process

Roman Rosipal

Department of Medical Cybernetics and Artificial Intelligence, Center for Brain Research, Medical University of Vienna, Freyung 6/2, A-1010 Vienna, Austria

roman.rosipal@meduniwien.ac.at

We propose and validate a continuous, entirely probabilistic model of the all night sleep process. The model is implemented as a hierarchical Gaussian Mixture Model (GMM). Features extracted from recordings following a polysomnographic setting are used. In the study we focus on describing sleep and transitions to sleep as a continuous process. The output of a GMM is a set of curves representing probability of each sleep or wakefulness state at a given time point. This is in contrast to the standard discrete Rechtschaffen & Kales (R&K) scoring system generally applied to score sleep in human subjects. The expected outcome of the new sleep modelling effort is additional information with respect to sleep quality, pathologies and other clinically relevant aspects not obtained by the R&K scoring. We validate the new sleep representation through a comparison with the R&K sleep profiles. We correlated the features extracted from both the discrete R&K and continuous GMM sleep profiles with external criteria of sleep represented by a set of psychometric variables collected after sleep. We demonstrate that the continuous sleep model possess the same or higher level of information about the sleep process and can successfully complement the R&K standard for a more comprehensive description of sleep.

Determinanty PZI ve zpracovatelském průmyslu ČR v letech 2000–2006

Eva Ryšavá

IES FSV UK, Opletalova 26, CZ–110 00 Praha 1

rysavaeva@volny.cz

Česká Republika je v posledních deseti letech úspěšným příjemcem zahraničního kapitálu. Proto je důležité porozumět chování těchto investorů a tomu, co hraje roli při jejich rozhodování, kam by měli umístit své investice. Cílem příspěvku je najít a odhadnout ekonometrický model determinant přímých zahraničních investic (PZI) ve zpracovatelském průmyslu v ČR v letech 2000–2006. Ekonometrický model zahrnuje několik základních ekonomických veličin. Vzhledem k tomu, že nelze zřejmě očekávat homogenní chování zahraničních investorů, je vedle jednoduchých technik použit i jeden z robustních odhadů.

Zjišťování optimálního počtu shluků ve statistických programových systémech

Hana Řezanková^[1], Dušan Húsek^[2]

^[1]VŠE, nám. W. Churchilla 4, CZ – 130 67 Praha 3; ^[2]ÚI AV ČR, Pod Vodárenskou Věží 2, CZ – 182 07 Praha 8

rezanka@vse.cz, dusan@cs.cas.cz

Príspevek se zaměřuje na možnosti vybraných statistických programových systémů v oblasti zjišťování optimálního počtu shluků při aplikaci metod shlukové analýzy. I když v literatuře byla pro stanovení optimálního počtu shluků navržena řada přístupů (z nejnovějších přehledových prací lze uvést [1] a [2]), jejich implementace v programových systémech poněkud zaostává. Na počet vhodných shluků se tak obvykle usuzuje pouze na základě grafického výstupu. Součástí metod rozkladu (k -průměrů) může být bodový graf pro první dvě hlavní komponenty (S-PLUS), součástí hierarchického shlukování bývá dendrogram. Spolu s ním lze zobrazit aglomerační schéma, v němž jsou pro každý krok shlukování uváděny vzdálenosti shluků, které byly právě spojeny (lze zkoumat difference vzdáleností). Pro hierarchickou shlukovou analýzu nabízejí další možnosti systémy SAS a SYSTAT (na základě vybraných indexů jsou zobrazovány grafy pro posloupnost počtu shluků do zadaného maxima – obvykle se hledá minimum či maximum na spojnici bodů). Pokud jde o metody rozkladu, pak v systému S-PLUS je pro metody k -medoidů (algoritmy PAM a CLARA) a pro fuzzy shlukovou analýzu (algoritmus FANNY) počítán obrysový koeficient (silhouette coefficient). Nejvhodnější je takový rozklad, pro který nabude tento koeficient nejvyšší hodnoty. Pro metody založené na hustotě poskytuje systém SAS neparametrický test pro počet shluků (stanovení počtu shluků na základě zadaných parametrů je ovšem součástí těchto metod, které jsou implementovány v systémech SAS a SYSTAT). Pro shlukování ve velkých souborech obsahujících proměnné různých typů je určena dvoukroková shluková analýza, která je implementována v systému SPSS. Její součástí je stanovení optimálního počtu shluků založené na informačních kritériích (Bayesovo a Akaikeovo). Možnosti programových systémů jsou ilustrovány na shlukování poslanců ruského parlamentu v roce 2004 na základě jednotlivých hlasování, viz [3] Kromě tradičních postupů, které jsou v systémech nabízeny, byla aplikována modifikace přístupu publikovaného v [4]. Jde o aplikaci faktorové analýzy na sledované objekty a následné shlukování objektů na základě získaných faktorových zátěží.

Reference

- [1] Abonyi J. a Feil B. (2007). *Cluster Analysis for Data Mining and System Identification*. Birkhäuser Verlag AG, Basel, Boston, Berlin.
- [2] Gan G., Ma C. a Wu J. (2007). *Data Clustering: Theory, Algorithms, and Applications*. ASA-SIAM, Philadelphia.
- [3] Húsek D., Frolov A.A., Polyakov P. a Řezanková H. (2006). *Neural Network Analysis of Russian Parliament Voting Patterns*. In: ISSA, G. (ed.). Computer

Science and Information Technology - CSIT 2006. : Applied Science Private University, Amman, 328–334.

- [4] Húšek D., Řezanková H. a Frolov A.A. (2006). *Overlapping clustering of binary variables*. In: Knowledge Extraction and Modeling [CD-ROM]. TILAPIA Edizioni, Italy, [http : //www.iasc - isi.org/Proceedings/2006/COMPSTAT_Satellites/KNEMO/](http://www.iasc-is.org/Proceedings/2006/COMPSTAT_Satellites/KNEMO/).

Pravděpodobnost ve středověku

Ivan Saxl

MFF UK, KPMS, Sokolovská 83, CZ–186 75 Praha 8

saxl@mat.cas.cz

Téměř obecně je přijímáno tvrzení, že teorie pravděpodobnosti vznikla v XVII. století při řešení *úlohy o rozdělení sázky* (rozdělení banku ve hře na N vítězných bodů přerušené za stavu $n : m; n, m < N$) B. Pascalem a P. de Fermatem v roce 1654. Úloha vznikla ve středověku a mezi její neúspěšné řešitele patří L. Pacioli, N.F. Tartaglia, G.F. Peverone a G. Cardano, který však zformuloval důležitý *princip úměrnosti*, podle něž řešení musí záviset pouze na počtech her, které hráčům k výhře chybějí.

V posledních letech však byly v italských rukopisech z první poloviny XV. století nalezeny (L. Toti Ricatelli v roce 1985 a R. Franci v roce 2003) dvě podobné úlohy, které vzbudily značnou pozornost, viz [1] a [2]. Jejich anonymní autoři (nazvaní N. Meusnierem *Ohri* – *Object historique relativement incertain* a *Ohrigens*) podali důmyslná aritmetická řešení a problém rozšířili i na případ účasti tří hráčů v přerušené hře, který byl také řešen v korespondenci Pascala a Fermata. Vedle rozborů jejich řešení je hledán původ podobných úloh v komerční oblasti, kde se pravděpodobnostní úvahy objevovaly v souvislosti s pojišťováním lodních nákladů a půjčkami na přepravované zboží, viz [3].

Za zmínku rovněž stojí blízkost Pascalových úvah ke starému talmudickému *principu dělení šatstva*, známému především z jeho aplikace na vysvětlení (po staletí hledaném) talmudem navrženého způsobu dělení dědictví, v němž se navíc uplatňuje také koaliční přístup.

Podobně jako v jiných oblastech historie přírodních věd se zde ukazuje problematičnost spojování počátků různých vědních oborů s konkrétními jednotlivci a úzce vymezenými daty.

Reference

- [1] Franklin J. (2001) *The Science of Conjecture: Evidence and Probability before Pascal*. The Johns Hopkins University Press, Baltimore.
- [2] Meusnier N. (2007) Le problème de partis bouge ...de plus en plus. *Journ@l Electronique d'Histoire des Probabilités et de la Statistique* **3**, 1–27, <http://www.jehps.net/>

- [3] Meusnier N. (2007) *Le problème de partis peut-il être d'origine arabomusulmane?* Journ@l Electronique d'Histoire des probabilités et de la Statistique **3**, 1–14, <http://www.jehps.net/>

Hypothesis testing by using N -distances

Bobosharif Shokirov

MFF UK, KPMS, Sokolovská 83, CZ–186 75 Praha 8

shokirov@karlin.mff.cuni.cz

In this talk we deal with two-sample hypothesis testing. We introduce a statistic $T(F_n, F_0)$ by N -metric and study its asymptotic distribution under the null and alternative hypothesis. Also we consider the criterion of *Asymptotic Relative Efficiency* and compare it with other criteria such as Kolmogorov or ω^2 .

Rank score test for independence with nuisance nonlinear regression

Martin Schindler

MFF UK, KPMS, Sokolovská 83, CZ–186 75 Praha 8 a

FP TUL, KAP, Studentská 2, CZ–461 17 Liberec

martin.schindler@matfyz.cz

We construct a class of tests for independence with nuisance nonlinear regression. The test statistic is based on the nonlinear regression rank scores. We illustrate the proposed test on the "Scottish hill races data - 2000" that gives record-winning times (male and female) in year 2000, distance and climb for 77 Scottish long distance races. We test the hypothesis of independence of the record-winning times and the record-winning times for females taking into account the influence of distance and climb on the record times (the form of the influence is given by nonlinear models).

Viacrozmerné testy o parametroch polohy viacerých súborov

Ján Somorčík

FMFI UK, KAMS, Mlynská dolina, SK–842 48 Bratislava

somorcik@fmph.uniba.sk

Ku klasickým nástrojom na testovanie hypotézy o rovnosti parametrov polohy viacerých mnohorozmerných rozdelení patrí Lawleyho-Hotellingov test, ktorý je veľmi kvalitný najmä v prípade normálne rozdelených dát. Ak však dáta pochádzajú z rozdelení s ťažkými chvostami, ukazuje sa ako rozumné nahradiť v jeho testovacej štatistike aritmetické priemery priestorovými mediánmi, ktoré sú robustnejšie.

Budeme sa zaoberať rôznymi verziami takto upraveného testu. Zaujímá nás budú najmä asymptotické vlastnosti, no pomocou Monte Carlo simulácií zhodnotíme aj konečnovýberové správanie v porovnaní s niektorými ďalšími testami.

Stochastický model vývoje epidemie s vakcináci

Jakub Staněk

MFF UK, KPMS, Sokolovská 83, CZ-186 75 Praha 8

stanekj@karlin.mff.cuni.cz

Tento príspevok sa zaoberá stochastickým modelom vývoje epidemie, ktorý je prirodzenou stochastickou verziou Kermack-McKendrickova modelu s vakcináci. Tato verzia je dána soustavou SDE

$$\begin{aligned} dX_t &= -\beta Y_t(X_t - \vartheta(Z_t))dt + \sqrt{\beta Y_t(X_t - \vartheta(Z_t))}dW_t, & X_0 = x_0 > 0 \\ dY_t &= \beta Y_t(X_t - \vartheta(Z_t))dt - \gamma Y_t dt - \sqrt{\beta Y_t(X_t - \vartheta(Z_t))}dW_t, & Y_0 = y_0 > 0 \\ dZ_t &= \gamma Y_t dt, & Z_0 = 0, \end{aligned}$$

kde X_t je počet zdravých jedinců, u kterých hrozí nákaza, Y_t je počet infikovaných jedinců, Z_t je počet imunních jedinců a $\vartheta(Z_t)$ je počet vakcinovaných jedinců.

V príspevku bude porovnané chování střední hodnoty procesů X, Y a Z s chováním těchto procesů v deterministickém Kermack-McKendrickově modelu, dále bude prezentována numerická optimalizace lineární vakcinace a nakonec bude diskutováno chování modelu v závislosti na velikosti populace.

Jak posuzovat spolehlivost softwaru

Jan Strouhal

ČVUT, CQR FS, Karlovo nám. 13, CZ-121 35 Praha 2

jstrouhal@email.cz

Klasifikační a regresní stromy v oblasti zvyšování spolehlivosti softwaru poskytují především nástroj pro zvýšení efektivity tohoto procesu. Vhodným výběrem modulů s vysokým očekávaným počtem chyb a odhadem jejich rizikovosti lze snížit náklady na testování. V literatuře jsou popsány nové vyvíjené metody, které usnadní a zlepší práci při tvorbě klasifikačního stromu. Jsou to metody rozšíření základní klasifikační škály na více skupin a metoda vyhledávání optimálních hranic dělení použitých metrik. Přínos těchto metod je ukázán na konečném příkladu. Pro snadnější aplikaci byl vytvořen program pro hledání optimálních kořenů metrik při konstrukci těchto klasifikačních stromů. Důležitými vstupními údaji jsou metriky, hodnoty metrik, moduly a třídy modulů. Uživatelé je umožněno získat mezivýsledky, jako jsou funkce výběru metrik F a konečný výsledek, kterým je nalezení minimálního E sloužícího k rozhodnutí o optimálním kořenu metriky. Nastavení programu je dostatečně variabilní. Program byl úspěšně testován na simulovaných i na skutečných datech.

Maximálně věrohodné odhady a lineární regrese ve výběrových šetřeních

Michaela Šedová, Michal Kulich

MFF UK, KPMS, Sokolovská 83, CZ-186 75 Praha 8

sedova@karlin.mff.cuni.cz

V klasické teorii výběrových šetření jsou předmětem studia parametry charakterizující konečnou populaci, jako např. úhrn nebo průměr N pevných hodnot. Příspěvek stručně popisuje situaci, kdy je vhodnější považovat pozorování za náhodné veličiny a zároveň je nutné brát v úvahu, že není k dispozici prostý náhodný výběr. Věnuje se výpočtu střední hodnoty v takovém případě, modifikuje metodu maximálně věrohodných odhadů a konkrétněji se zabývá lineárním modelem. Výsledky jsou ilustrovány na malé simulační studii.

Problém s příliš velkými daty – animal model

Petr Šimeček

VÚŽV, Přátelství 803, CZ-104 00 Praha 10 - Uhřetěves

simecek.petr@vuzv.cz

Animal model (AM) je statistický model biologie rutinně používaný při odhadu fenotypových hodnot jedince, plemenných hodnot, dědivosti... Z hlediska matematické statistiky se jedná o speciální případ modelu se smíšenými efekty

$$y = X\beta + Zu + e,$$

viz [1], kdy matice rozptylu náhodných efektů u je rovna $\sigma_u A$ pro A (známou) matici příbuznosti, viz [2].

V praxi se často setkáváme s případy, kdy je objem dat pro AM natolik velký, že jejich zpracování trvá týdny či vůbec není v reálném čase možné. V příspěvku se proto budeme zabývat možností odhadů v AM na základě opakovaného výběru vzorků z dat a stanovením chyby, které se tímto postupem dopouštíme.

Reference

- [1] Molenberghs G. (2001). *Linear Mixed Models for Longitudinal Data*. Springer, Berlin.
- [2] Henderson C.R. (1984) *Applications of Linear Models in Animal Breeding*. University of Guelph Press, Guelph, Canada.

Testy aditivity v analýze dvojného třídění bez opakování

Marie Šimečková

VÚŽV, Přátelství 803, CZ-104 00 Praha 10 - Uhřetěves

simeckova@gmail.com

Předpokládejme model

$$\mathbf{y} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \mathbf{e}_{ij}, \quad i = 1, \dots, a, \quad j = 1, \dots, b,$$

kde \mathbf{e}_{ij} jsou nezávislé normálně rozdělené náhodné veličiny s nulovou střední hodnotou a rozptylem σ^2 . Testovat významnost interakce γ_{ij} v tomto modelu není teoreticky obecně možné, ovšem přesto v praxi často nutné.

V literatuře bylo navrženo několik testů pro tento případ. V příspěvku jsou tyto výsledky shrnuty a zmíněno několik dalších možností. Testy jsou implementovány ve formě balíčku pro prostředí R `AdditivityTests`. V simulační studii jsou zkoumány vlastnosti těchto testů pro případ ANOVA modelu s jedním pevným a jedním náhodným faktorem.

Reference

- [1] Boik R.J. (1993). Testing additivity in two-way classifications with no replications: the locally best invariant test. *Journal of Applied Statistics* **20**, 41–55.
- [2] Alin A. a Kurt S. (2001). Testing non-additivity (interaction) in two-way ANOVA tables with no replication. *Statistical Methods in Medical Research* **15**, 63–85.

Mathematical models for epidemics

Josef Štěpán

MFF UK, KPMS, Sokolovská 83, CZ–186 75 Praha 8

`josef.stepan@mff.cuni.cz`

Kermack-McKendrick deterministic classics (1927) and its extensions. Particular and general stochastic models by means of stochastic differential equations. Partial differential equations and epidemics. Vaccinated epidemics.

Bayesian nonparametric estimation of hazard rate in survival analysis using Gibbs sampler

Jana Timková

MFF UK, KPMS, Sokolovská 83, CZ–186 75 Praha 8 a

ÚTIA AV ČR, Pod vodárenskou Věží 4, CZ–182 08 Praha 8

`timkova@karlin.mff.cuni.cz`

The contribution regards the nonparametric estimation of hazard rate in survival analysis as well as in more complicated schemes on real line. The technique of estimation is based on Bayesian approach and mild computational difficulties are overcome using the MCMC algorithms. The next step is expansion the approved method in 2D space. The technique is demonstrated in simulated data.

Nelineární statistická inference v oblasti nanotechnologií

Pavel Tuček^{*1}, Jiří Tuček², Jaroslav Marek³, Roman Procházka²

¹KGeo Přf UP, Třída Svobody 26, CZ – 771 46, Olomouc

²Centre for Nanomaterial Research, Šlechtitelů 11, CZ – 783 71 Olomouc

³KMAAM Přf UP, Tomkova 40, CZ – 779 00, Olomouc – Hejčín

pavel.tucek@upol.cz

Tento příspěvek je zaměřen na statistické metody a vývoj filtrovacích procedur v Mössbauerově spektroskopii. Statistické nástroje pro redukci šumu v měřeném spektru jsou užívány v mnoha oblastech. Šum v Mössbauerově spektroskopii je dán jako superpozice nerezonantních fotonů ve spojení s elektronickým šumem a rychlosti měřicího systému, která může být charakterizována jako vliv nelinearity zkoumaného procesu. Zcela nový matematicko-statistický přístup zlepšuje poměr signálu a šumu. Filtrovací procedura je založena na užití periodogramu a signifikantní hranice mezi šumem a užitečnými frekvencemi je zpětnovazební kontrolní mechanismus založený na korelační analýze. Finální fitování probíhá nejprve na základě dvouepochových modelů měření a nakonec za užití nelineárního modelu měření s podmínkou typu II. Korektnost tohoto řešení byla ověřena několika statistickými testy a bylo provedeno mnoho experimentálních měření s rostoucí kvalitou pro daný vzorek. Efekt této filtrovací procedury závisí na poměru užitečného signálu a šumu. Využitelnost této metody není univerzální, ale je vázána na některé specifické podmínky.

Adaptivní stochastické algoritmy v nelineární regresi

Josef Tvrdík

Ostravská universita, PřF, katedra informatiky, 30. dubna 22, CZ – 701 03 Ostrava 1

josef.tvrdik@osu.cz

Příspěvek se zabývá využitím adaptivních stochastických algoritmů v odhadu parametrů nelineárních regresních modelů. Adaptivní algoritmus řízeného náhodného prohledávání [3] je experimentálně porovnán s novými adaptivními verzemi diferenciální evoluce [1, 4] na úlohách z referenční databáze [2]. Uvažována je i možnost využití stochastických algoritmů k pohodlnému získání spolehlivých startovních hodnot pro standardní deterministické algoritmy.

Reference

- [1] Brest J., Greiner S., Boškovič B., Merník M., Žumer V. (2006) *Self-adapting control parameters in differential evolution: A comparative study on numerical benchmark problems*. IEEE Transactions on Evolutionary Computation **10**, 646–657.
- [2] NIST (2001) *Statistical reference datasets. Nonlinear regression*. NIST Information Technology Laboratory. <http://www.itl.nist.gov/div898/strd/>.

- [3] Tvrđík J., Křivý I., Mišík L. (2007) *Adaptive Population-based Search: Application to estimation of nonlinear regression parameters*, Computational Statistics and Data Analysis **52**, 713–724.
- [4] Tvrđík J. (2007) *Differential Evolution with Competitive Setting of its Control Parameters*. TASK Quarterly **11**, 169–179.

Metoda MLE pro logaritmicko-normální rozdělení s prahem

Vladimír Václavík

FAV ZČU, Univerzitní 22, 306 14 Plzeň, Czech Republic

vavra@kma.zcu.cz

Za odhad parametrů metodou maximální věrohodnosti (MLE) logaritmicko-normálního rozdělení se místo globálního maxima používá lokální maximum věrohodnostní funkce a tato metoda se většinou označuje LMLE. S nalezením lokálního maxima však mohou být spojeny obtíže. Předkládáme metodu, která odhady LMLE najde, pokud existují.

On efficiencies of decisions about statistical models based on f -divergences of empirical distributions

Igor Vajda

ÚTIA AV ČR, Pod vodárenskou Věží 4, CZ–182 08 Praha 8

vajda@utia.cas.cz

Statistical model of data source is a hypothetical probability distribution. True distribution of data is generally unknown - available are usually digitalized (appropriately quantized) i.i.d. data represented in a 1-1 way by empirical probability distribution. We consider statistical **decisions** consisting in the testing of validity of the hypothetical probability distributions.

Decision criteria studied: Values of the divergence goodness-of-fit statistics (f -divergences of hypothetical and empirical distributions) including the class of power divergence such as the classical Pearson statistics as well as the Neyman, the likelihood ratio and the Freeman-Tukey statistics.

Problem to be solved: Limit laws (asymptotic distributions) for the divergence statistics leading to the critical values for the asymptotically α -sized tests based on these statistics (i.e. to the tests with guaranteed decision errors of the first kind). and comparison of powers (decision errors of the second kind) of the tests based on various goodness-of-fit divergence statistics.

The first part of this paper deals with **limit laws** under hypotheses and local alternatives. New extension of the classical Morris (1975) limit law concerning the Pearson statistics to all divergence statistics. Innovative potential of the extension is demonstrated by a comparison with the classical result concerning the important

likelihood ratio statistic. The second part of the paper deals with **efficiencies** of the above considered decisions, i.e. with the powers of the tests based on various divergence statistics. It applies both the Pitman and the Bahadur approaches to the relative efficiency of statistics. All divergence statistics are shown to be equally efficient in the Pitman sense. However, some new results are presented demonstrating the maximal Bahadur efficiency of the likelihood ratio statistic in the important class of all power divergence statistics.

Některé vlastnosti váhové poloprostorové hloubky

Ondřej Vencálek

MFF UK, KPMS, Sokolovská 83, CZ-186 75 Praha 8

vencalek@karlin.mff.cuni.cz

Statistická hloubka dat patří k základním neparametrickým nástrojům pro analýzu mnohorozměrných dat. Nejběžnější způsob, jak definovat hloubku, je pomocí tzv. poloprostorové hloubky (halfspace depth, někdy též uváděn název Tukeyho hloubka). Zobecněním této hloubkové funkce je váhová poloprostorová hloubka. Tento příspěvek se zabývá vlastnostmi váhové poloprostorové hloubky v závislosti na volbě váhové funkce. Speciální pozornost je věnována souvislosti s nosičem rozdělení.

Consistency of weighted generalized method of moments

Jan Ámos Víšek

UK FSV, Smetanovo nábřeží 6, CZ-110 Praha 1

vissek@mbx.fsv.cuni.cz

In social sciences the underlying parametrized model (for a random “response” variable Y we are interested in) is frequently (over)identified by equations

$$\mathbb{E} [f(V, \beta^0)] = 0 \quad (1)$$

with random variables $V \in R^\ell \times \Omega$ and parameters $\beta^0 \in R^p$ to be estimated. Assume that all random variables are defined on a probability space (Ω, \mathcal{A}, P) and that the *true* value of parameters is vector β^0 . Without loss of generality we may assume that $\beta^0 = 0$. We frequently have $V = (Y, X)'$ with $Y \in R \times \Omega$ (*response variable*) and $X \in R^p \times \Omega$ (*explanatory variables*), but the structure of V can be generally a bit more complicated, including e. g. also some lagged values of variables in question.

(1) can be usually interpreted as a collection of orthogonality conditions since $f(V, \beta^0)$ has typically form

$$f(V, \beta^0) = U(V, \beta^0) \otimes Z(V, \beta^0) \quad (2)$$

where

$$u = U(V, \beta^0) \in R^k \times \Omega \quad \text{and} \quad z = Z(V, \beta^0) \in R^q \times \Omega \quad (3)$$

are *error term (disturbance)* and vector of the *(explanatory) variables* (or, if you want, of the *instruments*), respectively. The Kronecker product \otimes allows generally for possibility of $k > 1$, when we consider e. g. *seemingly unrelated regressions* or *simultaneous equations*, i. e. when the error term is multidimensional. The dependence of Z on its second argument is usually trivial, see also Hansen (1982).

If $p = q$, employing *method of moments*, for any data $\mathbf{V} = (V_1, V_2, \dots, V_n)'$ we can define $\hat{\beta}^{(MM,n)}$ as any solution of the empirical counterpart of (1), namely

$$g_n(\mathbf{V}, \beta) = \frac{1}{n} \sum_{i=1}^n f(V_i, \beta) = 0. \quad (4)$$

For the case when $p < q$, i. e. when number of orthogonal conditions is larger than dimension of parameter β , in 1982 Lars Peter Hansen proposed to consider an estimator by *generalized method of moments*

$$\hat{\beta}^{(GMM,n)} = \arg \min_{\beta \in R^p} [h_n(\mathbf{V}, \beta)]' \cdot h_n(\mathbf{V}, \beta) \quad (5)$$

where $h_n(\mathbf{V}, \beta) = a_n \cdot g_n(\mathbf{V}, \beta)$ and $\{a_n\}_{n=1}^\infty$ is a sequence of matrices of type $(p \times q)$ - generally with possibility of $a_n = a_n(\mathbf{V})$ - which he assumed to converge *almost surely* to a nonrandom matrix a . Then, considering $s = S(V, \beta) = a \cdot Z(V, \beta)$ and $f(V, \beta^0) = U(V, \beta^0) \cdot S(V, \beta^0)$, we can write (1) as

$$E [f(V, \beta^0)] = 0. \quad (6)$$

Now, for any $n \in \mathcal{N}$ let's consider $s_n = S_n(V, \beta) = a_n \cdot Z(V, \beta)$. Then, slightly modifying (2), indicating that the objective function can generally depend on n through the matrix a_n , we can define $f_n(V, \beta^0) = U(V, \beta^0) \cdot S_n(V, \beta^0)$ and put

$$g_n^{(GMM)}(\mathbf{V}, \beta) = \frac{1}{n} \sum_{i=1}^n f_n(V_i, \beta) = 0, \quad (7)$$

It shows that Hansen's proposal keeps in fact (4) for defining $\hat{\beta}^{(GMM,n)}$.

Now, notice please that (4) are normal equations for (5). It immediately indicates that atypical values among *residuals* $u_n = (U(V_1, \beta), U(V_2, \beta), \dots, U(V_n, \beta))'$ as well as among *instruments* $z_n = (Z(V_1), Z(V_2), \dots, Z(V_n))'$ can cause the same problems as in the case of regression model. The paper offers a proposal of robustified version of $g_n^{(GMM)}(\mathbf{V}, \beta)$ and a sketch of proof of its consistency.

Reference

- [1] Hansen, L. P. (1982): Large sample properties of generalized method of moments estimators. *Econometrica* **50**(4), 1029 – 1054.

Odhadování nerovnoměrnosti zatížení v paralelním systému

Petr Volf, Aleš Linka

ÚTIA AV ČR, Pod vodárenskou Věží 4, CZ – 182 08 Praha 8,
KHT, FT TUL, Studentská 2, CZ – 461 17 Liberec 1

volf@utia.cas.cz

Motivace může být třeba tato: Lano či textilní příze jsou složeny z více vláken, která ale nemusí být stejně napjatá. Při zatížení jsou tedy napínána a zatěžována nerovnoměrně, což vede k snížení odolnosti proti přetrhu. Pokud jsme schopni onu nerovnoměrnost zatížení popsat (nějakou náhodnou veličinou odchylek), je výsledná odolnost daná konvolucí. My se zabýváme opačnou úlohou - když výdrž systému neodpovídá rovnoměrnému sdílení zátěže, dá se odhadnout ona skrytá nerovnoměrnost, resp. zvlnění vláken v textilní přízi? Tohoto problému se bude týkat náš reálný příklad.

Ratio statistics – tolerance intervals

**Vokáčová Kateřina, Marek Patrice, Pavelka Tomáš,
Neumanová Martina, Šedivá Blanka, Vávra František**

FAV ZČU, Univerzitní 22, 306 14 Plzeň, Czech Republic

vavra@kma.zcu.cz

The results of independent experiments used for testing of some method or some phenomenon are often represented by samples $(r_1, s_1), (r_2, s_2), \dots, (r_n, s_n)$, where $r_i, s_i \geq 0$; $i = 1, \dots, n$, $\sum_{i=1}^n r_i > 0$. The main goal of our contribution is to analyse random variable $\xi_n = R/(R + S)$. The analysis concerns about tolerance bounds of statistic $\xi_n = (\sum_{i=1}^n r_i) / (\sum_{i=1}^n r_i + \sum_{i=1}^n s_i)$ and introduces its asymptotical model as well. Furthermore, there are examined two particular cases of n . In the first case, n is assumed to be non random nevertheless being changeable (e.g. comparison of digital voice recognition methods). In the other case, n is assumed to be fixed (e.g. calculation of a state unemployment rate from all district unemployment rates).

Přibližné konfidenční intervaly pre meranú hodnotu v prípade digitalizovaných meraní

Gejza Wimmer

MÚ SAV, Štefánikova 49, SK – 814 73 Bratislava

wimmer@mat.savba.sk

Uvažujeme problém konštruovania konfidenčného intervalu pre skutočnú meranú hodnotu v prípade digitalizovaných meraní. Predpokladáme, že chyby meraní sú normálne rozdelené, nezávislé, s nulovou strednou hodnotou a rovnakým neznámym rozptylom. Skutočné výsledky merania (realizácie normálnych náhodných veličín) sú v digitálnom tvare. Je to zapríčinené ohraňovanou (známou) rozlíšiteľnosťou δ meracieho prístroja, analogovo-digitálnym prevodom, atď. Keď disperzia chyby merania je veľká vzhľadom k δ , efekt digitalizácie zaniká a môžeme použiť štandardné štatistické postupy. Preto analyzujeme situácie, keď je rozptyl chyby merania relatívne malý.

Metódy výpočtu približných konfidenčných intervalov parametra polohy z digitalizovaných meraní

Gejza Wimmer Jr.

ÚM SAV, Dúbravská cesta 9, SK-841 04 Bratislava a MÚ SAV, Štefánikova 49, SK-814 73 Bratislava

wimmerg@mat.savba.sk

Príspevok sa zaoberá algoritmom pre výpočet približných konfidenčných intervalov parametra polohy z digitalizovaných meraní. Je prezentovaný všeobecný popis algoritmu pre generovanie náhodného výberu z fiduciálnej distribúcie ako aj podrobnejšie znázornenie vytvorenia množiny možných hodnôt parametrov (μ, σ) tak, aby táto bola neprázdna. Nasledujúcim krokom je vytvorenie výberového mechanizmu, ktorý vygeneruje jedinú hodnotu parametra (μ, σ) z polygónu všetkých možných hodnôt parametrov. Výsledky sú zhrnuté v simulačnej štúdii, ktorej cieľom bolo porovnať základné štatistické vlastnosti (pravdepodobnosť pokrytia a očakávanú dĺžku) približných konfidenčných intervalov pre parameter μ na základe malého výberu z digitalizovaných meraní.

Konfidenčné intervaly založené na digitalizovaných meraniach

Viktor Witkovský

ÚM SAV, Dúbravská cesta 9, SK-841 04 Bratislava

witkovsky@savba.sk

Uvedieme dve metódy konštrukcie približných konfidenčných intervalov pre parameter polohy popr. parametre lineárnej priamky, určených na základe nezávislých pozorovaní založených na digitalizovaných meraniach. Stručne spomenieme metódu založenú na teórii zovšeobecnenej fiduciálnej inferencie, ktorú nedávno navrhol Hannig (2006, 2008) ako zovšeobecnenie inferencie založenej na Fisherovej fiduciálnej pravdepodobnosti, pozri napr. Fisher (1935). Druhá metóda konštrukcie konfidenčných intervalov je založená na metóde invertovania asymptoticky korektného testu pomerom vierohodnosti, tzv. profile likelihood based confidence intervals.

Shlukování velkých souborů dat pomocí metod rozkladu

Marta Žambochová

FSE UJEP, KMS, Moskevská 54, CZ-400 96, Ústí nad Labem

zambochova@fse.ujep.cz

Příspěvek pojednává o shlukovacích metodách pro velké soubory dat. Zabývá se porovnáním několika variant vybraných metod rozkladu. Jedná se především o k-means algoritmus a jeho alternativy a dále metody využívající teorii centroidů. V některých variantách algoritmů jsou využity speciální typy stromů. Výhodou využití stromů v těchto metodách je fakt, že se v uzlech stromů shromáždí větší množství původních datových objektů a poté je možno s objekty jednoho uzlu pracovat vcelku, čímž se sníží objem zpracovávaných údajů.

Příhláška za člena České statistické společnosti

Osobní údaje	Adresa do zaměstnání
1.titul 2.titul Jméno Příjmení 3.titul e-mail	Organizace Útvar Ulice PSČ+město Telefon
Domácí adresa	
Ulice PSČ+město Telefon	
Odborné zájmy	Aplikace
matematická statistika výpočetní statistika ekonomická statistika státní statistika podniková statistika demografie ekonometrie jiné (slovy)	spol. vědách lékařství zemědělství techn. vědách jiné (slovy)

Zařazení v zaměstnání :

Další důležité informace jak se chcete případně zapojit do práce společnosti, další odborné zájmy atd.

Kam posílat další korespondenci (zakroužkujte): Domů Do práce

Podpis

Zašlete prosím na adresu: doc. RNDr. Jan Pícek, CSc., PF TUL, Studentská 2,
461 17 Liberec 1.



Generali Pojišťovna a.s. hledá do svého týmu kolegyni / kolegu na pozici pojistný matematik

Charakteristika pracovní činnosti

- pojistná matematika životního a neživotního pojištění
- analýza vývoje portfolia pojišťovny
- tvorba pravidelných i ad hoc reportů

Co požadujeme

- VŠ vzdělání matematického směru
- spolehlivost, samostatnost, iniciativu
- znalost anglického jazyka
- znalost MS Office (MS Excel na pokročilé úrovni)
- základy programování

Co nabízíme

- možnost spolupráce při studiu
- zázemí rostoucí pojišťovací skupiny
- možnost vzdělávání a odborného růstu
- přístup k celkovému pohledu na pojišťovnu
- kvalitní tým spolupracovníků
- široké spektrum zaměstnaneckých výhod

Kontakt:
Jiří Potužil
Generali Pojišťovna a. s.
Bělehradská 132, 12084 Praha 2
tel.: 221091402
email: jiri.potuzil@generali.cz

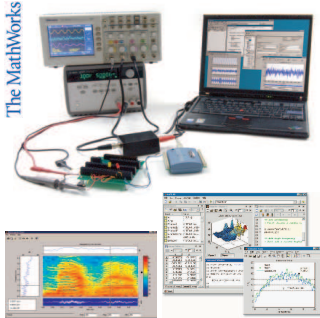
Pod křídly lva. 

Firma Humusoft vypsala jednu z cen
pro nejlepší vystoupení studentů
a doktorandů během ROBUSTu 2008

Univerzální výpočetní a simulační prostředí

MATLAB & Simulink

MATLAB je univerzální prostředí pro vědeckotechnické výpočty, modelování, návrhy algoritmů, simulace, analýzu a prezentaci dat, měření a zpracování signálů, návrhy řídicích systémů...

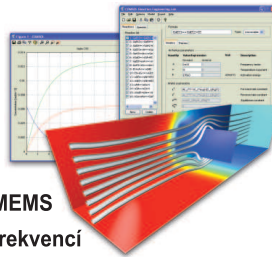


- měření, analýza a vizualizace dat
- vývoj a ověřování algoritmů
- návrh řídicích systémů
- zpracování signálů
- návrh komunikačních systémů
- konstrukční výpočty
- simulace dynamických systémů
- tvorba programových aplikací
- podpora systémů reálného času

COMSOL Multiphysics

COMSOL Multiphysics umožňuje řešit fyzikální úlohy popsané parciálními diferenciálními rovnicemi (PDE) metodou konečných prvků. Programem lze modelovat multifyzikální děje v inženýrské praxi a v mnoha vývojových oblastech technických a vědeckých oborů.


- simulace fyzikálních dějů metodou FEM
- otevřený, snadno ovladatelný systém
- knihovna vzorových příkladů
- komunikace s MATLABem
- dynamika tekutin, přestup tepla, akustika
- pružnost a pevnost, průmyslová chemie, MEMS
- elektromagnetismus nízkých a vysokých frekvencí
- průsak zeminy, těžba ropy, chemické reakce, ...



Výhradní distributor pro ČR a SR:

	www.humusoft.cz	info@humusoft.cz
	HUMUSOFT	Pobřežní 20 186 00 Praha 8, ČR

Firma Elkan vypsala jednu z cen
pro nejlepší vystoupení studentů
a doktorandů během ROBUSTu 2008



WOLFRAMRESEARCH
www.wolfram.com

COMPUTATION
DEVELOPMENT
DEPLOYMENT
VISUALISATION
MODELLING
SIMULATION
DOCUMENTATION

THINK
MATHEMATICA®

ONE INTEGRATED SOFTWARE SYSTEM DELIVERING
THE ULTIMATE TECHNICAL COMPUTING ENVIRONMENT

For further information please contact your local distributor:
Elkan, spol. s r.o. V Tůňích 12, 120 00 Praha 2, Czech Republic
PHONE +420-(0)2-24-999-111 FAX +420-(0)2-24-999-101
EMAIL vzak@elkan.cz WEB <http://www.elkan.cz>

*Alternatively, please contact
Wolfram Research directly:*
EMAIL info@wolfram.co.uk
PHONE +44(0) 1993 883400

© 2008 Wolfram Research, Inc. Mathematica is a registered trademark of Wolfram Research, Inc. All other trademarks are the property of their respective owners. Mathematica is not associated with Mathematica Policy Research, Inc. or Math Tech, Inc.