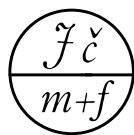
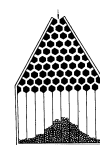


ROBUST 2016
Sborník abstraktů





ROBUST 2016 PODĚKOVÁNÍ



Organizátoři ROBUSTu 2016 by tímto chtěli poděkovat všem, kteří pomohli při přípravě a realizaci celé akce. Velmi důležitá pro nás byla morální i věcná podpora jak ČMS JČMF a ČStS, tak MFF UK a KPMS MFF UK. Neméně důležitá byla finanční podpora řady univerzit a sponsorů, ať již jmenovitých tak anonymních, která především umožnila účast mnoha studentů, jakož udělení cen za nejlepší vystoupení studentů a/nebo doktorandů.

Nejdůležitější pro zdárný průběh ROBUSTu však bylo úsilí všech účastníků, které věnovali jak přípravě a prezentaci svých vystoupení, tak vytvoření skutečně „robustní atmosféry“. K příjemné pohodě též přispěla velmi dobrá péče všech pracovníků Sporthotelu Kurzovní.

Všem děkují a na setkání na jubilejním dvacátém Robustu v zimě 2018 se těší
JA & GD & DH.



evropský
sociální
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání
pro konkurenceschopnost

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ



DataScripT

Antoch Jaromír	
<i>O detekci změn v panelových datech</i>	1
Běláček Jaromír	
<i>Prognóza demografických struktur pacientů ambulantně ošetřovaných v zařízeních Agel</i>	1
Brzezina Miroslav	
<i>Matematika v systémech pro určování polohy</i>	1
Burclová Katarína	
<i>Metódy odhadovania kovariančnej matice priestorového mediánu</i>	1
Cézová Eliška	
<i>Statistické vyhodnocování průmyslových dat</i>	2
Coufal David	
<i>Vzorkovací schémata v částicovém filtru</i>	2
Černý Michal	
<i>O jedné variantě lineární EIV regrese s omezenými chybami</i>	2
Čoupek Petr	
<i>Linear SDE's with additive noise of Volterra type</i>	3
Drabinová Adéla	
<i>Detection of differential item functioning with non-linear regression</i>	3
Dufek Jaroslav	
<i>Joint estimation of parameters of mortgage portfolio and the factor process</i>	3
Dvořák Jiří	
<i>Metoda minimálního kontrastu pro nehomogenní časoprostorové shlukové bodové procesy</i>	4
Fabián Zdeněk	
<i>Skalární skórová funkce</i>	4
Fačevicová Kamila	
<i>Souřadnicová reprezentace kompozičních tabulek</i>	4
Fišerová Eva	
<i>Regresní analýza kompozičních dat a její interpretace</i>	4
Gajdoš Andrej	
<i>Explicitný tvar momentov aproximácie MSE pre EBLUP v regresných modeloch časových radov</i>	5
Gardlo Alžběta	
<i>Imputace nulových hodnot v metabolomice</i>	5
Genest Christian, Nešlehová G. Johanna	
<i>Dependence modeling through copulas</i>	6
Hlubinka Daniel	
<i>Eliptické kvantily</i>	6
Holý Vladimír	
<i>Odhad integrované variance za přítomnosti mikrostrukturního šumu pomocí lineární regrese</i>	6
Horáková Hana	
<i>Detekce více změn v sezónním chování průtokových řad</i>	6
Houda Michal	
<i>Chance constrained DEA</i>	7
Hron Karel	
<i>Možnosti jednorozměrné statistické analýzy kompozičních dat</i>	7
Hrůzová Klára	
<i>Klasická a robustní ortogonální regrese mezi složkami kompozice</i>	8
Hudecová Šárka	
<i>Testy dobré shody pro časové řady s diskrétními veličinami</i>	8
Hušková Marie, Hlávka Zdeněk	
<i>Statistical procedures based on empirical characteristic functions</i>	8
Chvosteková Martina	
<i>Štatistická kalibrácia a tolerančné oblasti</i>	9
Jakubík Jozef	
<i>Výber regresorov v lineárnych zmiešaných modeloch s malým počtom prediktorov</i>	9
Janák Josef	
<i>Odhady parametrů v rovnici stochastického oscilátoru</i>	9
Jarušková Daniela	
<i>Nesimultánní změny ve složkách náhodného vektoru</i>	10
Jirsák Čeněk	
<i>Hledání optimálního řízení systému o dvou komponentách pomocí metody simulovaného žihání</i>	10

Jurczyk Tomáš	
<i>Robustifikace statistických a ekonometrických metod regrese</i>	10
Jurczyk Tomáš	
<i>Co se vyžaduje od moderních statistických programů</i>	11
Kadlec Karel	
<i>Ergodic Control for Lévy-driven linear stochastic equations in Hilbert spaces</i>	11
Kafková Silvie	
<i>Credibility premium in motor insurance</i>	11
Kasanický Ivan	
<i>Bayesova veta a asimilácia dát</i>	11
Kaspříková Nikola	
<i>Některé moderní přístupy k získávání dat</i>	12
Katina Stanislav	
<i>Analýza selhání ortopedických implantátů s využitím analýzy přežití a jádrového vyhlazování</i>	12
Kislinger Jan	
<i>Využití řízených markovských řetězců při optimalizaci cen jízdného ve vlaku</i>	12
Klaschka Jan	
<i>Za exaktní testy a konfidenční intervaly pro parametr binomického rozdělení logičtějši!</i>	12
Klebanov Lev	
<i>Big outliers versus heavy tails: What to use?</i>	13
Klicnarová Jana	
<i>Principy invariance pro náhodná pole</i>	13
Konečná Kateřina	
<i>Metoda maximální věrohodnosti pro volbu vyhlazovacích parametrů jádrových odhadů</i>	13
Koňasová Kateřina	
<i>Varianty K-funkce pro stacionární bodové procesy</i>	13
Kopa Miloš, Dupačová Jitka	
<i>Robustness in stochastic programs with decision dependent randomness</i>	13
Kopčová Veronika	
<i>Testing in the growth curve model</i>	14
Koščová Michaela	
<i>Parametrizované parciálne sumácie</i>	14
Kuruczová Daniela	
<i>Neparametrická funkcionálna regresia</i>	15
Lachout Petr	
<i>Diferencovatelnost reálných funkcí</i>	15
Maciak Matúš	
<i>Testing shape restrictions in LASSO regression</i>	15
Majdiš Mojmir	
<i>Lineárna kombinácia nezávislých náhodných premenných s lognormálnym rozdelením</i>	15
Martinková Patrícia	
<i>Flexibilní odhady reliability hodnocení v přijímacím řízení</i>	16
Masák Tomáš	
<i>Sparse principal component analysis</i>	16
Nagy Stislav	
<i>Hĺbka dát v konvexnej geometrii</i>	16
Navrátil Radim	
<i>Testy a odhady založené na minimalizaci vzdálenosti</i>	17
Nedela Roman	
<i>Sharp bounds on average treatment effects in the presence of sample selection bias</i>	17
Novák Petr	
<i>Diagnostické metody pro model zrychleného času</i>	17
Pawlas Zbyněk	
<i>Testování nezávislosti v prostorových modelech s kótami</i>	17
Pešta Michal	
<i>Mixed dynamic copulae for stochastic processes with application in insurance</i>	18
Pešková Barbora	
<i>Abrupt change in mean avoiding variance estimation and block bootstrap</i>	18
Petrová Barbora	
<i>Multidimensional stochastic dominance</i>	18

Picek Jan	
<i>L</i> -momenty s rušivou regresí	18
Pokora Ondřej	
Využití regrese a analýzy funkcionálních dat pro vyhodnocení neurofyzilogických záznamů	19
Prášková Zuzana	
Bootstrap pro závislá data a detekce změn	19
Rendlová Julie	
Analýza kategoriálních dat – problém vícenásobné volby v odpovědi	19
Rosa Samuel	
Optimal designs for dose-escalation studies	19
Rusá Šárka	
Bayesovská analýza tříúrovňového modelu mediace s ordinální odezvou	20
Selingerová Iveta	
Jádrové odhady jako alternativa (semi)parametrických modelů v analýze přežití	20
Sokol Ondřej	
Složitost výpočtu horní meze rozptylu nad náhodnými daty	20
Szűcs Gábor	
Rekurentné triedy diskretných rozdelení pravdepodobnosti a odhadovanie ich parametrov	21
Ševčovič Daniel	
Riccati transformation method for solving Hamilton-Jacobi-Bellman equation	21
Šimková Tereza	
Multivariate <i>L</i> -moment homogeneity test for spatially correlated data	21
Šulc Zdeněk	
Metodologie hodnocení měr podobnosti pro kategoriální data na velkém množství datových souborů ...	22
Talská Renata	
Kompoziční regrese s funkcí závisle proměnnou	22
Turčičová Marie	
Modelování kovariance v asimilaci dat	22
Vaňkátová Kristýna	
Shluková analýza ve směsích regresních modelů	23
Večeřa Jakub	
Estimation of parameters in a planar segment process model with a biological application	23
Vencálek Ondřej	
Zobecněné lineární modely nebo analýza kompozičních dat? Podobnosti a rozdílnosti!	23
Vinkler Mojmír	
Effect of denoising on brain atrophy measurements based on MRI for Alzheimer's disease	24
Víšek Jan Ámos	
Are the bad leverage points the most difficult problem for estimating regression model?	24
Volf Petr	
O analýze konkurujících si rizik s aplikací na čas prvního gólu v fotbalovém utkání	26
Witkovský Viktor	
Vybrané metody štatistickej inferencie založené na numerickej inverzii charakteristickej funkcie ...	26
Yermolenko Xeniya	
Non-unbiased two-sample nonparametric tests. Numerical example.	27
Zikmundová Markéta	
Procesy interagujících úseček	27

Jaromír Antoch**O detekci změn v panelových datech**

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

antoch@karlin.mff.cuni.cz

Uvažujeme model

$$y_{i,t} = \mathbf{x}_{i,t}^\top (\boldsymbol{\beta}_i + \boldsymbol{\delta}_i I\{t \geq t_0\}) + e_{i,t}, \quad 1 \leq i \leq N, 1 \leq t \leq T, \quad (1)$$

kde parametry modelu v i -tém panelu se změni v neznámém čase t_0 z $\boldsymbol{\beta}_i$ na $\boldsymbol{\beta}_i + \boldsymbol{\delta}_i$.

Hlavním cílem přednášky bude seznámit posluchače se zkušenostmi s odhadováním případné změny v tomto modelu při měnících se vstupních parametrech pro statistiky založené na procesu

$$U_N(t) = \sum_{i=1}^N (\hat{\boldsymbol{\beta}}_{i,t} - \hat{\boldsymbol{\beta}}_{i,T})^\top \mathbf{C}_{i,t} (\hat{\boldsymbol{\beta}}_{i,t} - \hat{\boldsymbol{\beta}}_{i,T}),$$

kde $\hat{\boldsymbol{\beta}}_{i,t}$ je odhad $\boldsymbol{\beta}$ získaný metodou nejmenších čtverců z prvních t pozorování a $\mathbf{C}_{i,t}$ je některá vhodná váhová matice.*Poděkování:* Jedná se o výsledky společné práce s M. Huškovou, J. Hanouskem a dalšími. Práce byla podpořena grantem GAČR P403/15/09663S.**Jaromír Běláček, Tomáš Fiala, Martin Parma, Pavel Michna, Karel Lukeš, Kateřina Murtingerová**
Prognóza demografických struktur pacientů ambulantně ošetřovaných ve zdravotnických zařízeních skupiny Agel

J.B. : Oddělení BioStat při ÚBI 1. LF UK Praha; VFN Praha

T.F. : Katedra demografie FIS VŠE

M.P. : Odbor plánování a controllingu AGEL, Prostějov

P.M., K.M. : AGEL Research a.s.; Ústav dějin medicíny a cizích jazyků, 1. LF UK Praha

Jaromir.belacek@vfn.cz

Předmětem našeho úsilí v r. 2015 bylo ověřit potenciál a možnosti propojení účelově vytříděných údajů z DB pacientů 13ti zdravotnických zařízení (ZZ) skupiny AGEL (nemocnice a polikliniky v kraji Moravskoslezském, Olomouckém a v Praze) s dostupnými údaji z demografických statistik ČR. Cílem tohoto shrnutí je identifikovat budoucí potřebu zdravotnické péče (podle nejvýznamnějších subkapitol číselníku MKN10) v porovnání s její spotřebou (monitorovanou v rámci speciálního číselníku zdravotnických odborností). V rámci každé dostupné nemocnice či polikliniky byly vytříděny počty pacientů podle unikátního RČ, roku, věku, pohlaví, ambulantních pracovišť a unikátních hlavních diagnóz. Aplikace projekčních koeficientů z demografické prognózy ČSÚ pro roky 2018, 2023 a 2028 vede k závěrům, že demografické stárnutí obyvatelstva bude mít zjevně ještě mnohem významnější odezvu na úrovni odpovídajících pohlavně-věkových struktur ambulantně ošetřovaných pacientů. Z výsledků formálních analýz vyplývá, které skupiny Dg a paralelně které skupiny zdravotnických odborností (na agregační úrovni) by měly nést v budoucnu největší zatížení z pohledu těch nejvíce exponovaných věkových skupin pacientů.

Poděkování: Práce na tomto příspěvku byly provedeny s podporou vedení společnosti AGEL Research, a.s.**Miroslav Brzezina****Matematika v systémech pro určování polohy**

PřF TUL, KAP, Studentská 2, 461 17 Liberec 1

miroslav.brzezina@tul.cz

V přednášce popíšeme řadu zajímavých matematických problémů spojených s problematikou určování polohy pomocí pomocí družic. Bude ukázáno, na jakých principech jsou tyto systémy založeny, jak je dosahováno velké přesnosti i to, jakou roli při tom hraje speciální i obecná teorie relativity.

Katarína Burclová, Ján Somorčík**Metódy odhadovania kovariančnej matice priestorového mediánu**

Univerzita Komenského, Mlynská dolina, 842 48 Bratislava

katarina.burclova@fmph.uniba.sk

Príspevok sa zaoberá odhadovaním kovariančnej matice priestorového mediánu, ktorý slúži ako robustný odhad parametra polohy viacrozmerých dát a je často interpretovaný ako optimálna poloha skladu v prípade, že máme

v priestore rozmiestnených niekoľko predajní, viď [2]. Simulačne sme porovnali vhodné metódy na odhad jeho kovariančnej matice - Bootstrap, Jackknife, Plug-In a metódu Boseho a Chaudhuriho [1]. Pre účely poslednej menovanej metódy sme najprv na základe simulácií určili optimálnu voľbu ladiacej konštantny nevyhnutnej pre výpočet. Totiž v [1] nie je otázka optimálnej voľby ladiaceho parametra riešená a ukázalo sa, že pri rôznych hodnotách tohto ladiaceho parametra vznikajú rôzne kvalitné odhady.

Literatúra

- [1] Bose, A., Chaudhuri, P.: On the dispersion of multivariate median. *Ann. Inst. Statist. Math.* **45**(3), 541–550 (1993)
- [2] Vardi, Y., Zhang, C.-H.: The multivariate L_1 -median and associated data depth. *The Proceedings of the National Academy of Sciences USA (PNAS)*. **97**(4), 1423–1426 (2000)

Eliška Cézová

Statistické vyhodnocování průmyslových dat

ČVUT FS, Technická 4, 166 07 Praha 6

eliskacqr@email.cz

Ve svém příspěvku se soustředím především na popis praktického využití popisné statistiky ve strojírenství při analýze kroutícího momentu potřebného pro utažení a povolení šroubového spoje pro dva typy závitů s různou povrchovou úpravou závitu matice.

David Coufal

Vzorkovací schémata v částicovém filtru

Ústav informatiky AV ČR, Pod Vodárenskou věží 2, 182 07 Praha 8

david.coufal@cs.cas.cz

Úloha filtrace spočívá ve stanovení optimálního odhadu nepozorovaných stavů náhodného procesu na základě pozorovaných dat. Matematicky se jedná o stanovení podmíněného rozdělení stavu procesu za podmínky dostupných pozorování. Toto podmíněné rozdělení se nazývá filtrační rozdělení. Úloha filtrace je řešitelná analyticky pouze ve speciálních případech, např. v případě lineárních Gaussovských procesů je řešením známý Kalmánův filtr. Obecně, ale tato úloha není v uzavřeném tvaru řešitelná a používají se aproximační algoritmy.

Částicový filtr představuje algoritmus, který sekvenčně generuje aproximaci filtračního rozdělení ve formě empirické míry, jejímž nosičem je soubor vzorků (částic) o daném rozsahu. Generování probíhá iteračně, kdy je v každém kroku generován nový soubor částic na základě předchozího souboru a aktuálního pozorování. Systém generování vzorků může mít různé varianty. V přednášce se zaměříme na tato různá vzorkovací schémata. Popíšeme, jak konkrétně vypadají a zmíníme jejich vlastnosti s ohledem na konvergenci empirické míry ke skutečnému filtračnímu rozdělení se vzrůstajícím počtem vzorků.

Michal Černý, Milan Hladík a Jaromír Antoch

O jedné variantě lineární EIV regrese s omezenými chybami

M.Č.: FIS VŠE, KE, Nám. W. Churchilla 4, 130 67 Praha 3

M.H.: MFF UK, KAM, Malostranské nám. 25, 118 00 Praha 1

J.A.: MFF UK, KPMS, Sokolovská 83, 186 75 Praha 8

cernym@vse.cz, hladik@kam.mff.cuni.cz, antoch@karlin.mff.cuni.cz

Regresní model je zatížen tzv. *EIV-problémem* (Errors-In-Variabes), jestliže namísto skutečných regresorů pozorujeme jen regresory zatížené chybami. Omezíme se na lineární regresní model, kde namísto matice regresorů (obecně stochastické) pozorujeme jen matici regresorů zatíženou aditivním stochastickým šumem (tzv. strukturální EIV model). Za jistých tradičních předpokladů lze regresní parametry „dobře“ odhadovat pomocí úplných nejmenších čtverců (Total Least Squares, TLS). Tato metodologie je také známa jako Demingova regrese. My se soustředíme na jiný, netradiční tvar chyb: budeme předpokládat, že všechny chybové distribuce mají omezený nosič v intervalu $(-\gamma, \gamma)$, kde $\gamma > 0$ (tzv. *poloměr chyb*) je neznámá konstanta. Doplníme další asymptotické předpoklady a sestrojíme konsistentní estimátor pro vektor regresních parametrů i pro poloměr chyb. Tento estimátor vznikne tak, že v TLS nahradíme Frobeniovu normu Čebyševovou normou. Z hlediska výpočetního stojí za zmínku, že výpočet estimátoru se redukuje na řešení systému zobecněných lineárně-fractionálních programů (generalized linear-fractional programming, GLFP), které lze efektivně počítat pomocí vnitřních bodů (Interior Point Methods, IPM). V příspěvku také ilustrujeme geometrii stojící za konstrukcí estimátoru.

Petr Čoupek

Linear SEE's with additive noise of Volterra type

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

coupek@karlin.mff.cuni.cz

In this talk, we consider the stochastic Cauchy problem (SCP)

$$\begin{aligned}dX_t &= AX_t dt + \Phi dB_t \\ X_0 &= x\end{aligned}$$

where B is a general, infinite-dimensional Volterra noise. The noise is continuous and satisfies a certain covariance structure determined by a Volterra kernel K but it does not need to be a semimartingale or a Gaussian process (e.g. the Rosenblatt process). We provide sufficient conditions for the existence and continuity of the mild solution to (SCP) with special attention to the interplay between the kernel K and the diffusion coefficient Φ .

Adéla Drabinová, Patricia Martinková

Detection of differential item functioning with non-linear regression

Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics, Charles University in Prague, Sokolovská 83, 186 75, Praha 8; Institute of Computer Science, The Czech Academy of Sciences, Pod Vodárenskou věží 271/2, 182 07, Praha 8

adela.drabinova@gmail.com; martinkova@cs.cas.cz

Detection of Differential Item Functioning (DIF) has been considered one of the most important topics in measurement. Procedure based on Logistic Regression is one of the most popular tools in study field, however, it does not take into account possibility of guessing, which is expectable especially in multiple-choice tests. In this work, we present an extension of Logistic regression procedure by including probability of guessing. This general method based on Non-Linear Regression (NLR) model is used for detection of uniform and non-uniform DIF in dichotomous items. NLR technique for DIF detection is compared to Logistic Regression procedure and methods based on three parameter Item Response Theory model (Lord's and Raju's statistics) in simulation study based on Graduate Management Admission Test. NLR method outperforms Logistic Regression procedure in power for case of uniform DIF detection and moreover by providing estimate of pseudo-guessing parameter. Proposed method also shows superiority in power at rejection rate lower than nominal value when compared to Lord's and Raju's methods. The proposed NLR method is accompanied by an R package difNLR and is implemented in an online Shiny application ShinyItemAnalysis.

Acknowledgement: Research was supported by the Czech Science Foundation project GJ15-15856Y.

Jaroslav Dufek, Martin Šmíd

Joint estimation of parameters of mortgage portfolio and the factor process

MFF UK, KPMS, Sokolovská 83, 186 75 Praha 8; ÚTIA AV ČR vii., Prague, The Czech Republic; ÚTIA AV ČR vii., Prague, The Czech Republic

jaroslav.dufek.2@seznam.cz; martinsmid.eu@gmail.com

In [1] a factor model for LGD (loss given default) and PD (probability of default) of mortgage portfolio based on KVM approach is proposed. The authors further fit an evolution of factors by a VECM model; however, they take the parameters of a portfolio as fixed instead of estimation. The present paper proposes a technique of a joint estimation of VECM and portfolio parameters in particular MLE function is defined; asymptotic properties are discussed. The present paper proposes a technique for joint estimation of the VECM and the portfolio parameters. In particular, MLE function is defined and its asymptotic properties are discussed. Finally, our technique is applied to US market data.

Literatúra

- [1] Gapko, P. and Šmíd, M.: Dynamic Multi-Factor Credit Risk Model with Fat-Tailed Factors. Czech Journal of Economics and Finance, 62(2): 125–140, 2012.

Jiří Dvořák

Metoda minimálního kontrastu pro nehomogenní časoprostorové shlukové bodové procesy

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

dvorak@karlin.mff.cuni.cz

Uvažujeme problém odhadu parametrů pro nehomogenní časoprostorové shlukové bodové procesy, s cílem vyhnout se obvyklým předpokladům na časoprostorovou separabilitu funkce intenzity. Zaměříme se na třídu shot-noise Coxových bodových procesů splňujících předpoklad intenzitou převáženou stacionaritou druhého řádu.

K odhadu využijeme metodu minimálního kontrastu na tzv. časoprostorovou K -funkci $K(r, t)$ jako funkci dvou argumentů: prostorové vzdálenosti dvou událostí r a jejich časové prodlevy t . Rozebereme postup odhadu využívající celou plochu $K(r, t)$ a také možnost zjednodušení postupu využitím pouze některých profilů této plochy.

Zdeněk Fabián

Skalární skórová funkce

ÚI AV ČR, Pod vodárenskou věží 2, 182 07 Praha 8

zdenek@cs.cas.cz

Skalární skórová funkce rozdělení F s hustotou f definovanou na celém R je funkce $-f'(x)/f(x)$. Autor již několikrát na konferencích ROBUST referoval o tom, že se mu tuto funkci s významem relativní vlivové funkce vzhledem k módu rozdělení podařilo zavést i pro ostatní spojitá rozdělení, definovaná na polopřímce nebo intervalu, s významem relativní vlivové funkce vzhledem k určitému centru rozdělení. V příspěvku novou funkci znovu definuji, vyšetřím jednoznačnost a upozorním na možné využití pro řešení statistických úloh.

Kamila Fačevicová, Karel Hron

Souřadnicová reprezentace kompozičních tabulek

PřF UPOL, KMAAM, 17. listopadu 12, 771 46 Olomouc

kamila.facevicova@gmail.com

Příspěvek je zaměřen na analýzu kompozičních tabulek, které představují přímé zobecnění D -složkových (vektorových) kompozičních dat. Kompoziční tabulky mohou být navíc chápány jako spojitá alternativa kontingenčních tabulek, také totiž zachycují vztah mezi dvěma faktory, založený na informaci o poměrech mezi prvky tabulky. Kvůli této relativní povaze se kompoziční tabulky (stejně jako kompoziční data obecně) řídí tzv. Aitchisonovou geometrií [1, 5]. Abychom tedy mohli použít standardní analytické metody, je potřeba tento typ dat převést prostřednictvím ortonormálních souřadnic do prostoru se standardní euklidovskou metrikou. Vyjádření v ortonormálních souřadnicích je běžně prováděno prostřednictvím tzv. postupného binárního dělení [2], takto získané souřadnice (balance) však nerespektují dvojrozměrnou povahu dat obsažených v kompozičních tabulkách. Kvůli zachování informace o vztahu mezi faktory proto navrhuje jejich doplnění o souřadnice, jejichž interpretace je úzce spjatá s poměry šancí mezi skupinami prvků [3]. Právě konstrukci těchto souřadnic a jejich interpretaci je věnována hlavní část příspěvku. Na případě čtyřpolní tabulky je navíc popsán vztah mezi navrhovanými souřadnicemi a parametry log-lineárního modelu.

Literatura

- [1] Aitchison J (1986) *The statistical analysis of compositional data*. Chapman and Hall, London.
- [2] Egozcue JJ, Pawlowsky-Glahn V, Mateu-Figueras G, Barceló-Vidal C (2003) *Isometric logratio transformations for compositional data analysis*. Math Geol 35:279–300.
- [3] Fačevicová K, Hron K, Todorov V, Templ M (2015) *Compositional tables analysis in coordinates*. Scandinavian Journal of Statistics. Přijato k tisku.
- [4] Pawlowsky-Glahn V, Egozcue JJ, Tolosana-Delgado R (2015) *Modeling and analysis of compositional data*. Wiley, Chichester.

Eva Fišerová, Ivo Müller, Karel Hron

Regresní analýza kompozičních dat a její interpretace

PřF UPOL, KMAAM, 17. listopadu 12, 771 46 Olomouc

eva.fiserova@upol.cz

Kompoziční data (kompozice) představují speciální typ mnohorozměrných dat, která v sobě nesou pouze relativní informaci. Výběrový prostor kompozičních dat je simplex, který spolu s tzv. Aitchisonou geometrií tvoří euklidovský vektorový prostor. Jelikož většina standardních statistických metod je založena na euklidovské geometrii v reálném prostoru, nelze kompoziční data přímo zpracovávat. Je třeba nejprve zvolit reprezentaci kompozičních dat v reálném prostoru pomocí vhodných transformací logaritmu podílů složek a až poté aplikovat standardní statistické metody. V přednášce se zaměříme na problematiku regresní analýzy kompozičních dat, kde mohou nastat tři základní úlohy: regrese s kompoziční vysvětlovanou proměnnou, regrese s kompozičními vysvětlujícími proměnnými a regrese mezi složkami kompozice. Cílem příspěvku je ukázat nové ortogonální souřadnice, které umožňují analogickou interpretaci regresních parametrů tak, jak je známa u standardní regrese.

Andrej Gajdoš, Martina Hančová

Explicitný tvar momentov aproximácie MSE pre EBLUP v lineárnych regresných modeloch časových radov

PF UPJŠ, Jesenná 5, 040 01 Košice 1

andrej.gajdos@student.upjs.sk

Jedným zo základných krokov empirickej predikcie časových radov pomocou najlepšieho lineárneho nevychýleného prediktora (tzv. EBLUP) je odhad jeho strednej štvorcovej chyby (MSE), ktorý nám dovoľuje vyjadriť napr. intervaly spoľahlivosti, či testovať hypotézy. Vo všeobecnosti analytický tvar zdokonaleného odhadu MSE pre EBLUP vhodný pre teoretické štúdium, ale aj počítačové simulácie, závisí od momentov (až šiesteho rádu) konečného pozorovania daného časového radu. Hlavným výsledkom príspevku je explicitný tvar daných momentov konečného pozorovania z viacrozmerného normálneho rozdelenia s nenulovou strednou hodnotou pre invariantných kvadratických odhadcov variančných parametrov časového radu. Praktickou aplikáciou výsledkov je ich použitie pre korekciu odhadu MSE v prípade, kde realizácie časového radu vedú k záporným štandardným odhadom (maximálna vierohodnosť alebo dvojitá metóda najmenších štvorcov) a kde je nutné použiť nezáporné invariantné vychýlené odhady. Dôležitosť tejto aplikácie sme podporili simulačnou štúdiou na reálnom ekonometrickom časovom rade modelovanom lineárnym regresným modelom, kde k spomenutým záporným odhadom dochádza v nezanedbateľnom množstve prípadov. Za prínos príspevku považujeme aj spôsob, akým boli dané explicitné tvary momentov získané. Išlo o aparát viacrozmernej štatistiky - vektorizácia, kroneckerov súčin, komutačné matice a vzťahy medzi nimi.

Podakovanie: Náš výskum je finančne podporený z grantov VEGA 1/0344/14 a VVGS 72616.

Alžběta Gardlo, Matthias Templ, Karel Hron, Peter Filzmoser

Imputace nulových hodnot v metabolomice

AG, KH: PrF UPOL, KMAAM, 17. listopadu 12, 771 46 Olomouc

AG: LF UPOL, Ústav molekulární a translační medicíny, Laboratoř metabolomiky, Olomouc

MT, PF: Dept. of Statistics and Probability Theory, Vienna University of Technology, Austria

alzbetagardlo@gmail.com

Soubor organických sloučenin, jejichž velikost je na úrovni molekul, které jsou obsaženy v daném biologickém materiálu, se nazývá metabolom. Jsou zde zahrnuty všechny organické látky přirozeně se vyskytující v metabolismu sledovaného živého organismu. Analýza metabolomu za daných podmínek se nazývá metabolomika. Při kvantifikaci informací z metabolomiky mají výsledky často podobu dat nesoucích pouze relativní informaci. Vektor těchto dat má kladné složky, relevantní informace je obsažena v podílech mezi nimi, případnou změnou měřítka se tedy tato informace nemění. Uvedená pozorování označujeme jako kompoziční data [1], jejich statistická analýza by měla uvedené vlastnosti zohledňovat.

Téměř žádná ze současných statistickým metod není schopna zpracovat datové soubory, které obsahují artefakty měření jako jsou chybějící hodnoty nebo tzv. hodnoty pod detekčním limitem. Speciálně tento druhý typ nulových hodnot se typicky vyskytuje v přírodních vědách jako je chemometrie (nebo specificky metabolomika) a je spojený s vlastnostmi měřícího přístroje, jehož limitace vedou právě k výskytu nulových hodnot. Tyto nulové hodnoty musejí být řádně nahrazeny před vlastní statistickou analýzou, založenou povětšinou na využití logaritmu původních látek či jejich podílů. Druhý z uvedených případů se pak týká právě kompozičních dat. Pro případ standardních mnohorozměrných souborů již existuje komplexní metodika imputace chybějících hodnot [2], kterou můžeme také aplikovat také na vysoce-rozměrná data [3], tato metodika ovšem selhává v případě kompozičních dat. Díky jejich specifickým vlastnostem musí být nahrazena každá hodnota s ohledem na její relativní charakter, vyjádřený pomocí podílů s ostatními složkami kompozice [4]. Prezentovaná metodika je navíc obohacena o přístup, který umožňuje imputaci nulových hodnot pro vysoce-rozměrná data. Teoretické aspekty jsou doplněny o simulační studii a reálný datový soubor z Laboratoře metabolomiky LF UP.

Literatura

- [1] Aitchison J. (1986). *The Statistical Analysis of Compositional Data*. Monographs on Statistics and Applied Probability. Chapman & Hall Ltd., London (UK). (Reprinted in 2003 with additional material by The Blackburn Press).
- [2] Little R.J.A., Rubin D.B. (2002). *Statistical analysis with missing data*. Wiley, Hoboken.
- [3] Walczak B., Massart D.L. (2001). Dealing with missing data. Part I. *Chemometrics and Intelligent Laboratory Systems* 58, 15–27.
- [4] Martín-Fernández J.A., Hron K., Templ M., Filzmoser P., Palarea-Albaladejo J. (2012). Model-based replacement of rounded zeros in compositional data: classical and robust approaches. *Computational Statistics and Data Analysis*, 56 (9), 2688–2704.

Christian Genest and Johanna Nešlehová

Dependence modeling through copulas

cgenest@math.mcgill.ca, johanna@math.mcgill.ca

Copulas are multivariate distributions whose margins are uniform on the interval $(0, 1)$. They provide a handy tool for the modeling of dependence between variables whose distributions are heterogeneous or involve covariates. Due to their flexibility, copula models are quickly gaining popularity in hydrology, finance and insurance. The first part of this two-hour talk will provide an introduction to statistical inference for copula models and its implementation in the R Project for Statistical Computing. In particular, it will be shown how estimation and goodness-of-fit testing can be performed using rank-based methods. The second part will be devoted to selected recent advances in the area, including copula modeling in the presence of ties, as well as the modeling of extreme-value dependence and its estimation through constrained B-spline smoothing.

Daniel Hlubinka

Eliptické kvantily

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

hlubinka@karlin.mff.cuni.cz

Představíme koncept eliptických kvantilů pro mnohorozměrná data, ukážeme možné rozšíření pro regresní model a pokusíme se nastínit použití eliptických kvantilů v analýze dat.

Vladimír Holý, Petra Tomanová

Odhad integrované variance za přítomnosti mikrostrukturního šumu pomocí lineární regrese

VŠE v Praze, Katedra ekonometrie, nám. Winstona Churchilla 1938/4, 130 67 Praha 3

vladimir.holy@vse.cz

Použití dat zaznamenaných o vysokých frekvencích je přínosné pro odhad integrované variance cen finančních aktiv. Při vysokých frekvencích se ovšem objevuje mikrostrukturní šum, který může značně vychýlit realizovaný rozptyl - základní odhad integrované variance. Navrhujeme alternativní odhad, který je robustní k mikrostrukturnímu šumu. Odhad využívá lineární regresi, ve které realizovaný rozptyl počítaný při různých frekvencích vystupuje jako vysvětlovaná proměnná a počet pozorování jako vysvětlující proměnná. Metodu lze využít i pro testování přítomnosti šumu v datech. Navrhovaný odhad se porovná s dalšími metodami popsány v literatuře na simulovaných datech pro několik modelů mikrostrukturního šumu.

Podpořeno z grantu IGS F4/63/2016 Vysoké školy ekonomické v Praze.

Hana Horáková

Detekce více změn v sezónním chování průtokových řad

FSv ČVUT, Katedra matematiky, Thákurova 7, 166 29 Praha 6

horakovah@mat.fsv.cvut.cz

V našich předchozích příspěvcích jsme se zabývali detekováním změn v sezónním chování průtokových řad. Roční chod byl popsán vektorem parametrů, které odpovídaly Fourierovým koeficientům v případě, že jsme roční chod aproximovali lineární kombinací sinů a kosinů, nebo koeficientům v metodě hlavních komponent. Pro takto zvolené vektory jsme použili statistické metody pro detekci bodu změny. V těchto metodách se předpokládá, že v řadě došlo maximálně k jedné změně (AMOC). Pokud v řadě došlo k více změnám, síla použitých testů se zmenšuje. Existují však i testy, které se používají v případě, že očekáváme v řadě více změn. Příspěvek se zabývá použitím těchto testů pro detekci změn sezónního chování průtokových řad.

Literatura

- [1] Antoch J., Jarušková D.: Testing for multiple change points. *Computational Statistics* 28, 2161–2184, 2013.
- [2] Jarušková, D., Horáková, H., Satrapa, L. Detection of non-stationarities of several small Czech rivers by statistical methods, *Civil Engineering Journal* 1, 2015, DOI: 10.14311/CEJ.2015.01.0005.

Poděkování: Příspěvek vznikl za finanční podpory SGS ČVUT SGS16/002/OHK1/1T/11.

Michal Houda

Chance constrained DEA

EF JČU, Branišovská 1645/31a, 370 05 České Budějovice

houda@ef.jcu.cz

We present a stochastic programming approach, namely that using joint probabilistic constraints, for a problem of evaluating the performance of decision-making units by data envelopment analysis (DEA). Inputs and outputs are considered random, described by their probability distributions. In such cases, the resulting optimization model can be formulated as chance constrained optimization problem; the constraints are required to be satisfied only up to a prescribed probability. We provide a deterministic reformulation of the set of feasible solutions based on this approach, report several properties of the model, computations challenges connected with the model and present a simple illustrative example.

Karel Hron

Možnosti jednorozměrné statistické analýzy kompozičních dat

PřF UPOL, KMAAM, 17. listopadu 12, 771 46 Olomouc

hronk@seznam.cz

Kompoziční data charakterizujeme jako mnohorozměrná pozorování, jejichž složky představují relativní příspěvky částí na celku; jinak řečeno, veškerá relevantní informace je obsažena v podílech mezi kompozičními složkami [1, 5, 5]. Typicky si lze kompozice představit jako proporcionalní data, ovšem podmínka konstantního součtu složek (v tomto případě 1) v definici kompozičních dat nehraje roli; rozhodnutí o tom, zda je datový soubor kompoziční či nikoli, souvisí právě s tím, zda je pro účely analýzy relevantní relativní struktura proměnných, či spíše jejich absolutní hodnoty. V kladném případě pak kompoziční data indukují tzv. Aitchisonovu geometrii (s vlastnostmi euklidovského vektorového prostoru), která respektuje jejich specifické vlastnosti, zejména zřejmou invarianci na změnu měřítka. Protože většina standardních mnohorozměrných statistických metod využívá (více či méně explicitně) předpokladu euklidovské geometrie v reálném prostoru, je potřeba před dalším statistickým zpracováním kompozice nejprve vyjádřit v interpretovatelných ortonormálních souřadnicích vzhledem k Aitchisonově geometrii [2]. Vzhledem k tomu, že je takovýchto souřadnic o jednu méně než je aktuální počet složek v kompozici, nelze reprezentovat všechny složky současně v rámci jednoho souřadnicového systému. Je ovšem možné zkonstruovat takový ortonormální souřadnicový systém, který v jediné souřadnici zachytí veškerou informaci o dané složce (ve smyslu agregace logaritmu podílů s touto složkou) [3, 6]. V rámci této základní myšlenky lze následně zohlednit vliv chyb měření pomocí vah složek, které se projeví v porušení exkluzivity jednorozměrného vyjádření relativní informace o dané kompoziční složce [4]. Teoretické úvahy budou doplněny simulačními studií a reálným příkladem z aplikací.

Literatura

- [1] J. Aitchison (1986). *The Statistical Analysis of Compositional Data*. Chapman & Hall, London.
- [2] J. J. Egozcue, V. Pawlowsky-Glahn (2005). Groups of parts and their balances in compositional data analysis. *Mathematical Geology*, 37(7), 795–828.
- [3] P. Filzmoser, K. Hron, C. Reimann (2009). Univariate statistical analysis of environmental (compositional) data: Problems and possibilities. *Science of the Total Environment*, 407(23), 6100–6108.
- [4] P. Filzmoser, K. Hron (2015). Robust coordinates for compositional data using weighted balances. In: K. Nordhausen, S. Taskinen (eds.) *Modern Nonparametric, Robust and Multivariate Methods*. Springer, Heidelberg: 167–184.
- [5] K. Hron (2010). Elementy statistické analýzy kompozičních dat. *Informační Bulletin ČStS*, 21(3), 41–48.
- [6] J.M. McKinley, K. Hron, E. Grunsky, C. Reimann, P. de Caritat, P. Filzmoser, K.G. van den Boogaart, R. Tolosana-Delgado (2016). The single component geochemical map: Fact or fiction. *Journal of Geochemical Exploration*, 162, 16–28.
- [7] V. Pawlowsky-Glahn, J.J. Egozcue, R. Tolosana-Delgado (2015). *Modeling and Analysis of Compositional Data*. Wiley, Chichester.

Klára Hružová¹, Valentin Todorov², Karel Hron¹, Peter Filzmoser³

Klasická a robustní ortogonální regrese mezi složkami kompozice

KH, KH: PřF UPOL, KMAAM, 17. listopadu 12, 771 46 Olomouc

VT: UNIDO, Vienna International Centre, Vienna, Austria

PF: Dept. of Statistics and Probability Theory, Vienna University of Technology, Austria

klara.hruzova@gmail.com

Při statistické analýze kompozičních dat [1] je třeba brát v úvahu jejich odlišné geometrické vlastnosti plynoucí z relativního charakteru kompozic, proto je nejprve potřeba tato data převést do souřadnic vzhledem k tzv. Aitchisonově geometrii. Základním problémem je, jaké souřadnice zvolit - záleží především na zvolené metodě a na související interpretovatelnosti výsledků [5]. Vhodná volba souřadnic se ukázala být důležitá zejména v případě, kdy se zajímáme o analýzu vztahu mezi jednotlivými složkami kompozice. Takto se dostáváme k regresnímu modelu, ve kterém jsou závisle i nezávisle proměnné zatíženy chybou [3], proto je jejich vztah modelován pomocí ortogonální regrese, založené na využití metody hlavních komponent [4]. Vzhledem k tomu, že reálná data často obsahují odlehlá pozorování, která mohou výsledky regresní analýzy znehodnotit, je kromě klasické ortogonální regrese aplikována i robustní verze využívající MM-odhady [6]. Statistická inference v regresním modelu (testování hypotéz o parametrech a konfidenční intervaly) je realizována pomocí neparametrického bootstrapu [2], v případě robustní metody pak využitím rychlého a robustního bootstrapu [7]. Teorie je představena na datech týkajících se hrubé přidáné hodnoty.

Literatura

- [1] Aitchison J. (1986). *The Statistical Analysis of Compositional Data*. Chapman & Hall, London.
- [2] Fox, J. (2002). *Bootstrapping Regression Models. Appendix to an R and S-PLUS Companion to Applied Regression*. <http://statweb.stanford.edu/tibs/sta305files/FoxOnBootstrappingRegInR.pdf>
- [3] Fuller, WA. (1987). *Measurement Error Models*. Wiley, New York.
- [4] Hružová K., Todorov V., Hron K., Filzmoser P. (2016). Classical and robust orthogonal regression between parts of compositional data. *Statistics*, DOI: 10.1080/02331888.2016.1162164.
- [5] V. Pawlowsky-Glahn, J.J. Egozcue, R. Tolosana-Delgado (2015). *Modeling and Analysis of Compositional Data*. Wiley, Chichester.
- [6] Rousseeuw P., Hubert M. (2013). High-breakdown estimators of multivariate location and scatter. In Becker C., Fried R., Kuhnt S. (eds.), *Robustness and complex data structures*, pp 49-66. Springer, Heidelberg.
- [7] Van Aelst S., Willems G. (2013). Fast and robust bootstrap for multivariate inference: The R package FRB. *Journal of Statistical Software* **53**(3).

Šárka Hudecová, Marie Hušková a Simon G. Meintanis

Testy dobré shody pro časové řady s diskrétními veličinami

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

hudecova@karlin.mff.cuni.cz

Modely časových řad diskrétních veličin (tj. modely posloupností závislých náhodných veličin s diskrétním rozdělením) nacházejí uplatnění v mnoha rozličných praktických situacích. Mezi nejpoužívanější patří modely celočíselné autoregrese (INAR) a modely celočíselné autoregrese s podmíněnou heteroskedasticitou (INARCH). V našem příspěvku se zabýváme testy dobré shody pro výše uvedené modely a navrhneme testovou statistiku založenou na empirické vytvořující funkci.

Marie Hušková, Zdeněk Hlávka

Statistical procedures based on empirical characteristic functions

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

huskova@karlin.mff.cuni.cz; hlavka@karlin.mff.cuni.cz

The empirical characteristic function (ECF) has been in use in statistical inference for more than forty years. It is known that there is a one-to-one relation between the distribution function (DF) and the characteristic function (CF). The talk will provide a partial overview of testing procedures based on the ECF within certain statistical models.

Specifically our emphasis is on recent developments of ECF procedures for goodness-of-fit testing, the two-sample and the k-sample problem, the change-point problem without and with nuisance parameters.

We discuss theoretical results (theorems and some proofs), computational aspects, simulations, and possible applications including change-point testing for the martingale difference hypothesis.

Martina Chvosteková

Štatistická kalibrácia a tolerančné oblasti

ÚM SAV, Dúbravská cesta 9, 841 04 Bratislava 4

chvosta@gmail.com

Vo viacerých vedných disciplínach sa často stretávame s úlohou stanovenia intervalového odhadu pre vysvetľujúcu premennú prislúchajúcu k pozorovanej hodnote vysvetľovanej premennej. Takáto úloha nastáva napríklad v prípadoch, keď určenie presnej hodnoty premennej pre subjekt je technicky (v lekárstve sa môže jednať aj o bolestivý invazívny zákrok pre pacienta), finančne, alebo aj časovo náročné, zatiaľ čo nameranie inej premennej pre daný subjekt, ktorá je lineárne závislá na hľadanej premennej, je nenáročné. Štatistická kalibrácia rieši úlohu určenia intervalového odhadu pre vysvetľujúcu premennú (kalibračný interval) prislúchajúcu k pozorovanej hodnote vysvetľovanej premennej na základe odhadnutého regresného modelu. Ak požadujeme, aby predpísaná časť skonštruovaných kalibračných intervalov odpovedajúca k ľubovoľnej postupnosti pozorovaných hodnôt vysvetľovanej premennej pokrývala prislúchajúcu skutočnú hodnotu vysvetľujúcej premennej, hovoríme o tzv. simultánnych kalibračných intervaloch (viacnásobne použiteľných konfidenčných intervaloch).

V príspevku sa budeme zaoberať konštrukciou jednorozmerných a viacrozmerých simultánnych kalibračných intervalov v prípade nezávislých normálne rozdelených pozorovaní, ktorá je úzko prepojená s problematikou konštrukcie simultánnych tolerančných oblastí. Tolerančné oblasti pokrývajú predpísanú časť rozdelenia s požadovanou spoľahlivosťou, v prípade lineárnej regresie však nie sú žiadne určené ako rovnomerne najužšie. Navrhnuté sú približné simultánne kalibračné oblasti za predpokladu rozdelenia vysvetľujúcich premenných, ktorých štatistické vlastnosti sú simulačne porovnané s vlastnosťami kalibračných oblastí skonštruovanými so známymi metódami.

Podakovanie: Táto práca bola podporená Agentúrou na podporu výskumu a vývoja APVV-15-0295.

Jozef Jakubík

Výber regresorov v lineárnych zmiešaných modeloch s malým počtom prediktorov

ÚM SAV, Dúbravská cesta 9, 841 04 Bratislava 4

jozef.jakubik.jefo@gmail.com

Uvažujeme lineárny zmiešaný model (LMM) v tvare

$$Y = X\beta + Zu + \varepsilon,$$

kde

Y je $n \times 1$ vektor pozorovaní,

X je $n \times p$ matica regresorov,

β je $p \times 1$ vektor neznámych pevných efektov,

Z je $n \times q$ matica prediktorov,

u je $q \times 1$ vektor neznámych náhodných efektov s rozdelením $\mathcal{N}(0, D(\theta))$, kde θ reprezentuje vektor variančných komponentov,

ε je $n \times 1$ vektor chýb s rozdelením $\mathcal{N}(0, R = \sigma^2 I)$ nezávislý od u .

Metódy na výber regresorov v LMM odvodené penalizáciou vierohodnostnej funkcie LMM pomocou ℓ_1 penalizácie vedú vo všeobecnosti na riešenie nekonvexného problému, z čoho vyplývajú limitácie na počet regresorov. V reálnych problémoch je ale často požiadavka na vysoký počet regresorov. V špecifickom prípade, keď je počet prediktorov malý, je ale postačujúce (ako ukážeme) uvažovať LMM ako lineárny regresný model s efektami s rôznou štruktúrou. To transformuje problém na konvexný, čo výrazne znižuje výpočtovú zložitosť problému. Predstavíme metódu na tom založenú, ktorá dáva porovnateľné výsledky pre problémy (s menším počtom regresorov) na ktoré sa dajú uplatniť doteraz známe metódy a zároveň dokáže riešiť problémy s väčším počtom regresorov. Metódu oprieme o konzistenciu zabezpečujúcu správny výber regresorov s rastúcim počtom pozorovaní ako aj o výsledky viacerých simulačných štúdií.

Podakovanie: Táto práca bola podporená Agentúrou na podporu výskumu a vývoja APVV-15-0295.

Josef Janák

Odhady parametrov v rovnici stochastického oscilátoru

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

janak@karlin.mff.cuni.cz

Uvažujme následující vlnovou rovnici

$$\begin{aligned}\frac{\partial^2 u}{\partial t^2}(t, \xi) &= b\Delta u(t, \xi) - 2a\frac{\partial u}{\partial t}(t, \xi) + Q^{\frac{1}{2}}\dot{B}(t, \xi), \quad (t, \xi) \in \mathbb{R}_+ \times D, \\ u(0, \xi) &= u_1(\xi), \quad \xi \in D, \\ \frac{\partial u}{\partial t}(0, \xi) &= u_2(\xi), \quad \xi \in D, \\ u(t, \xi) &= 0, \quad (t, \xi) \in \mathbb{R}_+ \times \partial D,\end{aligned}$$

kde $D \subset \mathbb{R}^d$ je otevřená omezená množina s hladkou hranicí, $a > 0$ a $b > 0$ jsou neznámé parametry, $Q^{\frac{1}{2}}$ je pozitivní nukleární operátor v $L^2(D)$ a $\dot{B}(t, \xi)$ je formální časová derivace Brownova pohybu závislého na prostorové proměnné.

Na základě pozorování trajektorie procesu $X^T = \{X_t = (u(t, \cdot), \frac{\partial u}{\partial t}(t, \cdot))^*, 0 \leq t \leq T\}$ pořídíme silně konzistentní odhady parametrů a a b .

Daniela Jarušková

Nesimultánní změny ve složkách náhodného vektoru

FSv ČVUT, KM, Thákurova 7, 166 29 Praha 6

daniela.jaruskova@cvut.cz

V časových okamžicích $i = 1, \dots, n$ pozorujeme posloupnost nezávislých dvojrozměrných vektorů $\{(X_1(i), X_2(i))\}$ takových, že korelační koeficient $\text{corr}(X_1(i), X_2(i)) = \rho$ je znám.

Uvažujeme dva statistické problémy. Za prvé je třeba rozhodnout, zda došlo ke změně ve střední hodnotě buď $\{X_1(i)\}$ nebo $\{X_2(i)\}$, případně obou posloupností, přičemž však předpokládáme, že ke změně nemuselo nutně dojít ve stejný časový okamžik.

Pokud dojdeme k závěru, že ke změně došlo, chceme časové okamžiky, v kterých ke změně došlo, odhadnout a najít jejich přibližné (asymptotické) rozdělení.

Čeněk Jirsák

Hledání optimálního řízení systému o dvou komponentách pomocí metody simulovaného žhání

FP TUL, KAP, Univerzitní náměstí 1410/1, 461 17 Liberec

cenek.jirsak@tul.cz

Príspevek se zabývá optimálním řízením dvojice spojitě zastarávajících komponent. Konkrétně použitím algoritmu simulovaného žhání na optimalizaci řízení systému. Obě komponenty mají společně za úkol dodat požadovaný konstantní výkon. Komponenty zastarávají podle toho, jak jsou používány. Ve chvíli, kdy jsou obě komponenty natolik opotřeбенé, že nevládnou dodat požadovaný výkon, jedna z nich (případně obě) se vymění za novou. Optimální je v tomto případě takové řízení systému, které na dlouhodobém horizontu minimalizuje počet výměn komponent za jednotku času.

Motivace pochází z průmyslové aplikace optimalizace údržby. Konkrétně jde o skupinu komponent dodávající společný výkon, jako jsou například mlýny na uhlí v tepelné elektrárně, případně filtry v úpravně pitné vody. Uvažovaný model je zatím jednoduchý a slouží jako základ k dalšímu zkoumání problematiky.

Tomáš Jurczyk

Robustifikace statistických a ekonometrických metod regrese

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

jurczyk@karlin.mff.cuni.cz

Dva z problémů, které se mohou vyskytnout během regresní analýzy, jsou multikolinearita regresorů a přítomnost odlehklých pozorování. Tento příspěvek se snaží vyšetřit a vysvětlit chování regresních metod v situacích, kdy se v datech vyskytují oba tyto problémy zároveň. Ukazuje se, že klasické robustní metody mohou mít v těchto případech problémy. Z tohoto důvodu byla navržena nová metoda, která se tyto problémy snaží řešit. Budou odvozeny její vlastnosti, popíšeme její chování a využití jako diagnostického nástroje.

Tomáš Jurczyk

Co se vyžaduje od moderních statistických programů

Dell Computers, V Parku 2325/16, 148 00, Praha 4

tomas.jurczyk@software.dell.com

V této přednášce si povíme o tom, že ideální software pro analýzu dat již dávno není jen o nutnosti mít implementovány rozličné analytické metody. Ukážeme si přímo na ukázce v softwaru, které vlastnosti by měl moderní software mít, aby usnadnil a urychlil práci kolektivu pracovníků, kteří potřebují s daty a modely pracovat, nebo těch, kteří se na základě dat potřebují rychle rozhodovat. S tím souvisí také potřeby velkých institucí (jako jsou například banky). Na závěr nakousneme, jakým směrem se platformy pro analyzování dat vyvíjejí a na co je aktuálně kladen důraz ve vývoji (IoT, In-database analytics, Collective intelligence, Edge analytics, ...).

Karel Kadlec

Ergodic Control for Lévy-driven linear stochastic equations in Hilbert spaces

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

Karel.Kadlec.9@seznam.cz

In this contribution, controlled linear stochastic evolution equations driven by cylindrical Lévy processes are studied in the Hilbert space setting. The control operator may be unbounded which makes the results obtained in the abstract setting applicable to parabolic SPDEs with boundary or point control. The first part contains some preliminary technical results, notably a version of Itô formula which is applicable to weak/mild solutions of controlled equations. In the second part, the ergodic control problem is solved: The feedback form of the optimal control and the formula for the optimal cost are found. As examples, various parabolic type controlled SPDEs are studied.

Silvie Kafková

Credibility premium in motor insurance

Comenius University, Mlynska dolina, 842 48 Bratislava, Slovakia

175424@mail.muni.cz

In 2005 a price war among insurance companies started in the Czech Republic. The number of contracts of motor insurance increased but the premium was reduced. Consequently, the motor insurance became unprofitable. In this paper, the problem of poor using of rating system in the vehicle insurance is solved. Mathematical approaches are used for it, namely Bayesian correction according to the credibility theory. A priori insurance premium depends on annual expected claim frequency, which is modelled by Poisson regression from observable risk factors. However, the existence of unobservable risk factors results in risk level heterogeneity within tariff groups. To remove heterogeneity we propose individual insurance premium calculation by Bayesian techniques of premium corrections. Then the fairer premium for drivers will be obtained and loss of this product will be eliminated.

Ivan Kasanický

Bayesova veta a asimilácia dát

ÚI AV ČR, Pod Vodárenskou věží 2, 182 07 Praha 8

kasanicky.ivan@gmail.com

Základným predpokladom zmysluplného použitia Bayesovej vety pre asimiláciu dát je splnenie podmienky

$$\int_{\mathcal{H}} \exp\left(-\frac{1}{2}\langle x-y, R^{-1}(x-y)\rangle\right) d\mu(x) > 0 \quad (2)$$

kde \mathcal{H} je separabilný Hilbertov priestor, μ je Gaussova miera na tomto priestore, R je ľubovoľný kovariančný operátor a $y \in \mathcal{H}$.

Splnenie tejto podmienky je očividné v prípade, že dimenzia stavového priestoru \mathcal{H} je konečná. V opačnom prípade, keď $\dim(\mathcal{H}) = \infty$, nie je splnenie podmienky (2) vôbec jasné a ukážeme si príklady, keď je vyššie uvedený integrál rovný nule. Navyše si ukážeme, že ak je operátor R kovarianciou nejakej merateľnej náhodnej veličiny definovanej na \mathcal{H} , tak podmienka (2) nie je splnená pre skoro všetky $y \in \mathcal{H}$.

Na druhej strane si ukážeme, že ak prisputíme aby data boli iba slabá náhodná veličina, tak podmienka (2) je splnená pre veľkú triedu kovariančných operátorov R .

Nikola Kaspříková

Některé moderní přístupy k získávání dat

VŠE v Praze, Katedra matematiky, nám. W. Churchilla 1938/4, 130 67 Praha 3
school@tulipany.cz

Díky rozvoji webových služeb a dalších datových technologií se podstatně rozšiřují možnosti analýz. To klade nároky na některé složky procesu analýzy dat – vedle vlastního statistického modelování je důležité taky zvládnutí činností, jako jsou vyhledání dat, získání a příprava dat, vizualizace a (v současnosti už skoro nezbytně interaktivní) prezentace výsledků. Některé novější datové technologie, které se v poslední době v praxi prosazují, krátce představím a v duchu populární Data Science se pokusíme jemně nahlédnout do nedávné historie Robustů.

Stanislav Katina, Iveta Selingerová, Andrea Kraus, Ivana Horová, Jiří Zelinka

Analýza selhání ortopedických implantátů s využitím analýzy přežití a jádrového vyhlazování

PřF MU, ÚMS, Kotlářská 2, 611 37 Brno
katina@math.muni.cz

Cílem našeho příspěvku je představit analýzu datového souboru 46 859 operací kyčelního kloubu získaného ze Slovenského artroplastického registru (SAR). Tyto operace byly realizovány ve všech 40 ortopedických a traumatologických klinikách ve Slovenské republice. Zajímali jsme se především o dobu do prvního selhání jednotlivých implantátů a faktory, které by tuto dobu mohly ovlivnit. Soubor obsahoval údaje od 1. ledna 2003 do 31. prosince 2014, maximální doba sledování pak byla 12 let. Z celkového počtu operací došlo ve 1005 případech k selhání implantátu a následné revizi. Vzhledem k přítomnosti cenzorovaných dat jsme využili metody analýzy přežití, zejména pak rizikovou funkci, která vyjadřuje okamžitou pravděpodobnost selhání implantátu. Při modelování rizikové funkce jsme aplikovali metody jádrového vyhlazování. Kromě klasické rizikové funkce jsme se zaměřili také na podmíněnou rizikovou funkci, kde sledovanou kovariátou byl věk. Data jsem stratifikovali podle diagnózy, typu fixace a pohlaví.

Jan Kislínger

Využití řízených markovských řetězců při optimalizaci cen jízdného ve vlaku

MFF UK, KPMS, Sokolovská 83, 186 75 Praha 8
jan.kislínger@gmail.com

Jedná se o modifikaci známé úlohy *Newsboy problem*. V tomto případě je však množství zásob (počet sedadel ve vlaku) pevně dané a rozhodovací proměnnou jsou ceny jízdenek na jednotlivých trasách. Jedná se tedy o problém s endogenní náhodností, jelikož poptávka po jízdenkách je ovlivněna jejich cenou. Další vlastností tohoto modelu je fakt, že existuje konkurence na straně poptávky i v případě, že cestující chtějí koupit jízdenky na jiné trasy. Skutečná kapacita sedadel pro jednotlivé trasy je tedy náhodná a závisí na poptávkách po jízdenkách na jiné trasy.

Předpokládáme, že aktuální obsazenost vlaku (množství prodaných jízdenek) se řídí nehomogenním markovským řetězcem. Intenzity přechodu tak závisí na zvolené ceně. Řízení markovského řetězce tak probíhá skrze volbu ceny jízdného.

Jan Klaschka

Za exaktní testy a konfidenční intervaly pro parametr binomického rozdělení logičtěší!

ÚI AV ČR, Pod Vodárenskou věží 2, 182 07 Praha 8
klaschka@cs.cas.cz

Od 50. let 20. století byla navržena řada méně konzervativních exaktních alternativ ke klasickému Clopper-Pearsonovu konfidenčnímu intervalu pro parametr binomického rozdělení (interval Sterneho, Blyth-Still-Casellův, Blakerův, ...) a odpovídajících testů. Tyto alternativní metody ovšem mají mnohdy různé nežádoucí vlastnosti a vyžadují úpravy, které by je uvedly do souladu se „zdravým rozumem“. Přednáška bude pojednávat o pěti typech nelogického chování: (i) Nesouvislé konfidenční množiny. (ii) Konfidenční meze nemonotónní vzhledem k hladině spolehlivosti (non-nestedness). (iii) Rozpor mezi testem a konfidenčním intervalem při chybné volbě jejich typů. (iv) Rozpory, mezi testem a konfidenčním intervalem zdánlivě (ale pouze zdánlivě) nevyhnutelné i při správné volbě jejich typů. (v) Rozpory mezi inferencemi při různě velkém počtu pokusů. Možnosti, jak těmto nelogičnostem čelit, proberu z metodologického i výpočetního hlediska. Mimo jiné stručně shrnu příspěvky z ROBUSTů 2010, 2012 a 2014 a budu referovat o novějších výsledcích spolupráce s kolegy Jenő Reiczigelem (HU) a Månsem Thulinem (SE).

Lev B. Klebanov

Big outliers versus heavy tails: What to use?

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

levkl@gmail.com

The notions of outliers and tails are often used in different senses. Our aim in the talk is to analyze possible strong mathematical definitions of both that notions, study their connection between each other, and consider their possible use in applications.

Jana Klicnarová

Principy invariance pro náhodná pole

EF JČU, Studentská 13, 370 05 České Budějovice

klicnarová@ef.jcu.cz

V tomto příspěvku se budeme věnovat centrálním limitním větám a principům invariance pro slabě závislá náhodná pole. Nejprve si připomeneme problematiku limitních vět pro náhodná pole využívajících aproximací (pomocí martingalů či m -závislých náhodných polí). Poté představíme a porovnáme některé základní výsledky, kterých bylo v této oblasti dosaženo.

Kateřina Konečná, Ivana Horová

Metoda maximální věrohodnosti pro volbu vyhlazovacích parametrů jádrových odhadů podmíněné hustoty

PřF MU, ÚMS, Kotlářská 2, 611 37 Brno

xkonecn3@math.muni.cz; horova@math.muni.cz

Technika jádrového vyhlazování patří mezi neparametrické metody a je vhodným nástrojem pro odhad podmíněné hustoty. Jádrové odhady závisí na volbě jádrové funkce, která hraje roli vah, a na šířce vyhlazovacích parametrů. Právě vyhlazovací parametry mají největší vliv na kvalitu výsledného odhadu. Při špatné volbě šířky vyhlazovacích parametrů může dojít k přílišnému podhlazení nebo naopak přehlazení odhadu.

V tomto příspěvku se budeme zabývat metodami pro odhad šířky vyhlazovacích parametrů. Navrhovaná metoda maximální věrohodnosti bude porovnána s dobře známou metodou křížového ověřování pomocí simulační studie. Použití obou metod bude rovněž doplněno o aplikaci na reálném datovém souboru.

Literatura

- [1] Bashtannyk, D. M., Hyndmann, R. J. (2001). *Bandwidth Selection for Kernel Conditional Density Estimation*. Computational Statistics & Data Analysis **36**(3), 279–298
- [2] Fan, J., Yim, T. H. (2004). *A crossvalidation Method for Estimating Conditional Densities*. Biometrika **91**(4), 819–834

Poděkování: Tento příspěvek byl podpořen Grantovou agenturou ČR (grant GA15-06991S).

Kateřina Koňasová

Varianty K-funkce pro stacionární bodové procesy

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

konasova.k@seznam.cz

K -funkce představuje cenný nástroj pro posuzování regularity nebo naopak tendence k vytváření shluků u bodových procesů, sloužících k modelování bodových vzorů v rovině nebo prostoru. Varianty této popisné charakteristiky využíváme při směrové analýze bodových procesů. Na reálných i simulovaných datech je zde ilustrováno použití tzv. směrové K -funkce při detekci dominantního směru v bodových vzorech.

Miloš Kopa, Jitka Dupačová

Robustness in stochastic programs with decision dependent randomness

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

kopa@karlin.mff.cuni.cz

Results of stochastic optimization problems are often influenced by the model misspecification and simplifications, or by errors due to approximations, estimations, and incomplete information. The obtained optimal solutions, recommendations for a decision maker, should be then carefully analyzed. We shall deal with output analysis, robustness, and stress testing with respect to uncertainty or perturbations of input data for risk constrained portfolio optimization problems via the contamination technique and the worst-case analysis. We focus on problems with decision dependent random returns. Applying the contamination techniques we present lower and upper bounds for optimal value function for several different decision dependent randomness problems.

Veronika Kopčová

Testing in the growth curve model

University of Pavol Jozef Šafárik, Košice, Slovakia

veronika.kopcova@student.upjs.sk

In this work we consider the testing of intraclass structure of the covariance matrix in the growth curve model. This model is of the form $Y = XBZ + \varepsilon$, $\text{var}(\text{vec}(\varepsilon)) = \Sigma \otimes I$, $E(\varepsilon) = 0$ where Y is matrix of observations. We solve the problem of testing hypothesis $H_0 : \Sigma = \sigma_1 I + \sigma_2 \mathbf{1}\mathbf{1}'$, $B = 0$ against $H_1 : \Sigma > 0$, $B \neq 0$ using likelihood ratio test procedure.

Literature

- [1] Khatri, C. G. (1973). Testing some covariance structure under s growth curve model. *Journal of multivariate analysis*, Vol. 3, 102–116.
- [2] Kollo, T. and D. von Rosen (2005). Advanced multivariate statistics with matrices. *Springer Netherlands*, ISBN: 978-1-4020-3418-3

Acknowledgement The support of the grant VEGA MŠSR 1/0344/14 and VVGS-PF-2016-72616 is kindly announced.

Michaela Koščová, Ján Mačutek, Gejza Wimmer

Parametrizované parciálne sumácie

MK, JM: FMFI UK, KAMŠ, Mlynská dolina, 842 48 Bratislava

GW: MÚ SAV, Štefánikova 49, 814 73 Bratislava

michaela.koscova@fmph.uniba.sk

Uvažujeme parciálnu sumáciu

$$P_x(a) = \sum_{j=x}^{\infty} g(j)P_j^*(a), \quad x = 0, 1, 2, \dots, \quad (3)$$

kde $\{P_j^*(a)\}_{j=0}^{\infty}$ je diskrétné rozdelenie pravdepodobnosti, ktoré nazývame rodič a $\{P_j(a)\}_{j=0}^{\infty}$ je diskrétné rozdelenie pravdepodobnosti, ďalej nazývané potomok (pozri [1, 2]), obe definované na nezáporných celých číslach, a je parameter. Obmedzíme sa len na jednoparametrické diskrétné rozdelenia pravdepodobnosti. V [1] je uvedená nutná a postačujúca podmienka pre invarianciu vzhľadom na parciálnu sumáciu (3),

$$g(j) = 1 - \frac{P_j^*(a)}{P_{j+1}^*(a)}. \quad (4)$$

Táto podmienka je splnená práve vtedy, keď sa rodič aplikáciou parciálnej sumácie (3) nezmení, teda $P_x^* = P_x$, $x = 0, 1, 2, \dots$. Pre zdôraznenie úlohy parametra môžeme písať $g(x) = g(x, a)$.

Teraz uvažujeme modifikáciu parciálnej sumácie (3)

$$P_x = c \sum_{j=x}^{\infty} g(j, \lambda)P_j^*(a), \quad x = 0, 1, 2, \dots, \quad (5)$$

kde je zachovaný vzťah (4), avšak hodnota parametra a je nahradená inou hodnotou λ , c je vhodná konštanta, taká, aby $\{P_j\}_{j=0}^{\infty}$ bolo rozdelenie pravdepodobnosti (teda aby súčet jednotlivých pravdepodobností bol 1).

Rozdelenie $\{P_j\}_{j=0}^{\infty}$ buď závisí od dvoch parametrov, λ a a , pre $a \neq \lambda$, alebo druhý parameter λ je eliminovaný vďaka normovacej konštante c . Podľa toho, ktorá z týchto dvoch možností nastáva, je možné kategorizovať všetky jednoparametrické diskrétné rozdelenia pravdepodobnosti. Napríklad, aplikáciou parciálnej sumácie (5) na Poissonovo rozdelenie (rodič) vzniká potomok - nové rozdelenie s dvoma parametrami, takže Poissonovo rozdelenie zaradujeme medzi rozdelenia citlivé na zmenu parametra parciálnej sumácie. Naopak, napríklad geometrické rozdelenie patrí medzi rozdelenia rezistentné voči zmene parametra parciálnej sumácie, teda medzi rozdelenia, ktoré po aplikácii parciálnej sumácie (5) zostávajú nezmenené.

Literatúra

- [1] Mačutek, J. (2003). On two types of partial summation. *Tatra Mountains Mathematical Publications*, **26**, 403–410.
- [2] Wimmer, G., Mačutek, J. (2012). New integrated view at partial-sums distributions. *Tatra Mountains Mathematical Publications*, **51**, 183–190.

Podakovanie: Podporené grantom VEGA 2/0047/15 (M. Koščová, J. Mačutek, G. Wimmer) a grantom UK/138/2016 (M. Koščová).

Daniela Kuruczová, Jan Koláček

Neparametrická funkcionálna regresia

PrF MU, ÚMS, Kotlářská 2, 611 37 Brno

369088@mail.muni.cz

Tento príspevok predstavuje neparametrickú jadrovú regresiu pre funkcionálne dáta ako rozšírenie klasickej jadrovej regresie. Sústreďme sa hlavne na problematiku voľby vyhladzovacieho parametra a testovanie metódy na reálnych dátach.

Poděkování: Tato práce byla podpořena grantem GAČR GA15-06991S.

Petr Lachout

Diferencovatelnost reálných funkcí

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

lachout@karlin.mff.cuni.cz

Základní úlohu v teorii optimalizace hraje diferencovatelnost reálných funkcí a její různá zobecnění; viz [1], [2], [3]. Příspěvek zavede a představí pojem diferencovatelnosti a druhé diferencovatelnosti funkce více proměnných z pohledu teorie optimalizace. Dále je pokusem o upřesnění a jednotné zavedení značení.

Literatura

- [1] Bazara, M.S.; Sherali, H.D.; Shetty, C.M.: *Nonlinear Programming. Theory and Algorithms*. 2nd edition, Wiley, New York, 1993.
- [2] Rockafellar, T.: *Convex Analysis*, Springer-Verlag, Berlin, 1975.
- [3] Rockafellar, T.; Wets, R. J.-B.: *Variational Analysis*. Springer-Verlag, Berlin, 1998.

Matúš Maciak

Testing shape restrictions in LASSO regression

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

maciak@karlin.mff.cuni.cz

The LASSO regularized regression approaches are suitable in situations where one needs to estimate the unknown model itself and, at the same time, choose relevant covariates from a much larger set of hypothetical regressors. We propose an effective estimation algorithm based on the L_1 -norm regularization which is motivated by some machine learning ideas and techniques. Moreover, using some small modifications we are also able to incorporate various shape restrictions and change-points into the final model and taking advantage of some recent post-selection results we are also able to verify whether the imposed shape constraints (change-points respectively) are statistically relevant for the model or not. We propose a statistical test which can be used to derive a proper decision. We also investigate some finite sample properties via a simulation study and a real data example.

Mojmír Majdiš

Lineárna kombinácia nezávislých náhodných premenných s lognormálnym rozdelením

ÚM SAV, Dúbravská cesta 9, 841 04 Bratislava

mojmir.majdis@savba.sk

Lognormálne rozdelenie a lineárna kombinácia takto rozdelených nezávislých náhodných premenných majú široké využitie v technike, prírodných vedách alebo ekonómii. Napriek tomu stále chýbajú použiteľné balíčky

v štatistickom softvéri (napríklad R), ktoré dokážu vypočítať kvantilovú funkciu (QF), hustotu (PDF) alebo distribučnú funkciu rozdelenia (CDF) s vysokou presnosťou. Uvádzame numerické výpočty týchto funkcií s použitím numerického invertovania charakteristickej funkcie (CF).

Najprv sa musíme vysporiadať s problémami, ktoré vznikajú pri samotnom výpočte hodnôt charakteristickej funkcie v numerických softvéroch - veľa štandardne uvádzaných foriem CF je vo forme nekonečného radu ktorý v týchto softvéroch nekonverguje, prípadne konverguje príliš pomaly. Potom invertujeme charakteristickú funkciu na PDF (alebo CDF, QF) použitím algoritmu rýchlej Fourierovej transformácie (FFT). Nakoniec interpolujeme získanú hustotu a distribučnú funkciu na požadované body - najčastejšie na požadované body kvantilovej funkcie.

Náš algoritmus taktiež testujeme na niekoľkých príkladoch na ilustráciu využitých metód a taktiež porovnávame výsledky získané z týchto príkladov so známymi výsledkami (ak také existujú).

Podakovanie: Táto práca bola podporená Agentúrou na podporu výskumu a vývoja APVV-15-0295.

Patrícia Martinková

Flexibilní odhady reliability hodnocení v přijímacím řízení s využitím smíšených lineárních modelů

ÚI AV ČR, Pod Vodárenskou věží 2, 182 07 Praha 8

martinkova@cs.cas.cz

Reliabilita, tedy spolehlivost hodnocení během přijímacího řízení má velký vliv na to, do jaké míry je toto hodnocení schopné predikovat kvalitu budoucího studenta či zaměstnance. Instituci se tedy může vyplatit zaměřit svou snahu na to, aby byla hodnocení dostatečně spolehlivá, např. aby hodnocení dvou různých hodnotitelů byla co nejvíce konzistentní. V příspěvku si ukážeme, jak lze konzistenci hodnotitelů (tzv. inter-rater reliability) odhadovat v komplexních designech a jak lze testovat hypotézy o tom, zda je inter-rater reliability ovlivněna např. typem uchazeče nebo zkušeností hodnotitele.

Poděkování: Výzkum je podporován grantem GA ČR GJ15-15856Y.

Tomáš Masák

Sparse principal component analysis

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

tom.masak@gmail.com

Principal component analysis (PCA) is a well-known dimensionality reduction technique which, however, behaves poorly when the number of variables is comparable to, or larger than, the number of observations. Sparsity assumption is usually utilized in such setting, leading to sparse PCA, which is known to be NP-hard. In this contribution, we discuss a regression-based approach to sparse PCA and we present an iterative reweighted least squares (IRLS) algorithm for sparse PCA. An extensive simulation study is carried out to show that our IRLS algorithm has a superior performance over the original regression-based approach. Our method share many favorable properties with the regression-based approach and, moreover, it surpasses abilities of the regression-based approach to correctly identify important variables, to explain variance in data and to produce estimates of principal components that are close to their population counterparts.

Stanislav Nagy

Hĺbka dát v konvexnej geometrii

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

KU Leuven, Statistics Section, Celestijnenlaan 200b, 3001 Leuven, Belgium

nagy@karlin.mff.cuni.cz

Od počiatku 90. rokov sa polopriestorová hĺbka stala hojne diskutovaným, ale tiež pomerne exotickým a v rámci teórie izolovaným, nástrojom neparаметrickej analýzy mnohorozmerných dát. V prehľadovom príspevku poodhalíme vzťahy medzi konceptom hĺbky dát používaným výlučne v štatistike, a niektorými pojmami známymi v konvexnej a diferenciálnej geometrii. Ukážeme, že v obore konvexnej geometrie existujú dôležité matematické výsledky priamo aplikovateľné na problémy vyvstávajúce pri skúmaní teórie polopriestorovej hĺbky. S ich pomocou (čiastočne) vyriešime niektoré odolávajúce problémy týkajúce sa hĺbky dát, a naznačíme možnosti budúceho výskumu a nových aplikácií tak pre hĺbku, ako aj pre afinne invariantnú geometriu.

Radim Navrátil

Testy a odhady založené na minimalizaci vzdálenosti

PřF MU, ÚMS, Kotlářská 2, 611 37 Brno

navratil@math.muni.cz

Je známo, že odhad metodou nejmenších čtverců v klasickém lineárním regresním modelu je velmi citlivý na porušení předpokladů modelu, zejména normality chyb. Proto byla zavedena celá řada odhadů, které tyto nedostatky dokázaly překonat. Velice zajímavou třídu těchto odhadů tvoří tzv. R-odhady, které místo původních pozorování pracují pouze s jejich pořadími. Cílem toho příspěvku bude tuto třídu rozšířit o další odhady založené pouze na pořadích na základě minimalizace jistých vzdáleností. Ukážeme, že v některých situacích mají tyto odhady větší vydatnost než klasické odhady. Zaměříme se také na testování hypotéz o regresním parametru založené na předchozích odhadech.

Roman Nedela, Lukáš Lafférs

Sharp bounds on average treatment effects in the presence of sample selection bias and survey non-response using linear programming approach

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

roman.nedela3@umb.sk; lukas.laffers@umb.sk

This paper reformulates the problems of bounding average treatment effects under sample selection bias and survey non-response studied in [1] and [2] as linear programs. This allows researchers to conduct sensitivity analysis of identifying assumptions easily while the bounds remain sharp. We provide a mathematical formulation of the problems, replicate existing analytical results and extend them into sensitivity analysis.

Literature

- [1] Lee, David S. *Training, wages, and sample selection: Estimating sharp bounds on treatment effects*, The Review of Economic Studies 76.3 (2009): 1071-1102.
- [2] Behaghel, Luc, Crépon B., Gurgand M., Le Barbanchon T. *Please Call Again: Correcting Nonresponse Bias in Treatment Effect Models*, The Review of Economics and Statistics 97.5 (2015): 1070-1080.

Petr Novák

Diagnostické metody pro model zrychleného času

FIT ČVUT, KAM, Thákurova 9, 160 00 Praha 6

petr.novak@fit.cvut.cz

Regresní model zrychleného času umožňuje interpretovat závislost doby přežití jedince na vysvětlujících proměnných za přítomnosti censorování. V příspěvku zkoumáme diagnostické metody pro tento model a rozšiřujeme možnosti testování dobré shody modelu s daty. Využíváme jak postupy založené na reziduích lineární regrese modifikovaných pro censorovaná data, tak teorii čítacích procesů. Na příkladech zkoumáme vlastnosti popsanych metod v různých situacích.

Zbyněk Pawlas

Testování nezávislosti v prostorových modelech s kótami

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

pawlas@karlin.mff.cuni.cz

Prostorové bodové procesy představují užitečné modely pro analýzu náhodně rozmístěných bodů v prostoru. Tyto body odpovídají polohám sledovaných událostí. Často máme k dispozici i dodatečnou informaci spojenou s každým bodem. Tu zahrnujeme do modelu v podobě tzv. kót a mluvíme o kótovaných bodových procesech. Nejjednodušší situace nastává, když jsou kóty nezávislé na polohách, pak můžeme obě složky modelovat zvlášť. Důležitým úkolem je proto otestovat hypotézu nezávislosti poloh a kót na základě pozorovaných dat. V přednášce představíme některé vhodné postupy. Bodové procesy slouží rovněž jako základní kámen pro budování složitějších geometrických modelů jako jsou některé náhodné uzavřené množiny. K množině představující oblasti, ve kterých se projevuje sledovaná událost, může být také přidána dodatečná informace v podobě kóty. Ta je v tomto případě náhodnou funkcí na dané náhodné množině. Opět nás může zajímat test nezávislosti těchto dvou náhodných složek.

Michal Pešta

Mixed dynamic copulae for stochastic processes with application in insurance

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

pesta@karlin.mff.cuni.cz

The theoretical goal is to study mathematical properties of the mixed dynamic copulae, i.e., copulae having continuous as well as discrete margins and are time-varying. Moreover, we discuss their extensions for the conditional and hierarchical case.

The practical aim is to apply the derived theory for loss reserving in non-life insurance. An insurance company puts sufficient provisions from the premium payments aside, so that it is able to settle all the claims (losses) that are caused by these insurance contracts. The main issue is how to determine or estimate these claims reserves, which should be held by the insurer so as to be able to meet all future claims arising from policies currently in force and policies written in the past.

Barbora Peštová

Abrupt change in mean avoiding variance estimation and block bootstrap

ÚI AV ČR, Pod Vodárenskou věží 2, 182 07 Praha 8

pestova@cs.cas.cz

We deal with sequences of weakly dependent observations that are naturally ordered in time. Their constant mean is possibly subject to change at most once at some unknown time point. The aim is to test whether such an unknown change has occurred or not. The change point methods presented here rely on ratio type statistics based on maxima of cumulative sums. These detection procedures for the abrupt change in mean are also robustified by considering a general score function. The main advantage of the proposed approach is that the variance of the observations neither has to be known nor estimated. The asymptotic distribution of the test statistic under the no change null hypothesis is derived and is free of any tuning parameters. Moreover, we prove the consistency of the test under the alternative. A block bootstrap method is developed in order to improve computational performance of the asymptotic methods. The validity of the bootstrap algorithm is shown. The results are illustrated through a simulation study.

Barbora Petrová

Multidimensional stochastic domin

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

petrova@karlin.mff.cuni.cz

Stochastic dominance is a form of stochastic ordering, which stems from decision theory when one gamble can be ranked as superior to another one for a broad class of decision makers whose utility functions, representing preferences, have very general form. There exists extensive theory concerning one dimensional stochastic dominance of different orders. However it is not obvious how to extend the concept to multiple dimension which is especially crucial when utilizing multidimensional non separable utility functions. One possible approach is to transform multidimensional random vector to one dimensional random variable and put equivalent stochastic dominance in multiple dimension to stochastic dominance of transformed vectors in one dimension. We suggest more general framework which does not require reduction of dimensions of random vectors in order to define stochastic dominance so as to be able to employ multidimensional non separable utility functions in the considerations. We seek for a generator of stochastic dominance of considered orders in terms of von Neumann – Morgenstern utility functions. Moreover, we develop necessary and sufficient conditions for one random vector to stochastically dominate another one, similarly as it is derived for one dimensional case. We focus on behavior of such defined conditions for a specific choices of utility functions or a specific choice of multivariate distribution of random vectors.

Jan Píček, Martin Schindler

L-momenty s rušivou regresí

FP TUL, KAP, Studentská 2, 461 17 Liberec 1

jan.picek@tul.cz

L-moment je analogie obvyklého momentu a má podobnou interpretaci. Je definován jako lineární kombinace středních hodnot pořádkových statistik. Hosking a Balakrishan (2015) ukázali, že L-momenty jsou speciálním

případem L-odhadů. V příspěvku navrhuje zobecnění L-momentů v modelu lineární regrese založené na regresních kvantilech jako speciální L-odhad. Vlastnosti rozšířených L-momentů jsou ilustrovány na simulovaných datech.

Ondřej Pokora, Jan Koláček

Využití regrese a analýzy funkcionálních dat pro vyhodnocení neurofyziologických záznamů

PřF MU, ÚMS, Kotlářská 2, 611 37 Brno

pokora@math.muni.cz; kolacek@math.muni.cz

Příspěvek se zabývá statistickou analýzou neuronálních záznamů vyvolané sluchové aktivity u krysy, kde je aktivita sluchových neuronů sledována jako vyvolané potenciály v závislosti na různých zvucích, hlasitostech a podané látce. Data byla původně zpracována s využitím nelineární regrese v závislosti na hlasitosti zvuku, přičemž odhadnuté parametry byly pro jednotlivé zvuky porovnány, [1]. V současnosti pro vyhodnocení záznamů využíváme metody analýzy funkcionálních dat, kdy jsou měření chápána jako body na křivkách. Nástroji statistické analýzy jsou zde jádrové vyhlazování, funkcionální PCA, funkcionální regrese a klasifikace. V příspěvku se zmíníme o obou použitých přístupech a dosažených výsledcích.

Literatura

- [1] Wan, I., Pokora, O., Chiu, T.W., Lansky, P. and Poon, P.W. (2015) Altered intensity coding in the salicylate-overdose animal model of tinnitus. *BioSystems* 136, 113–119.

Poděkování: Tato práce byla podpořena grantem GAČR GA15-06991S.

Zuzana Prášková

Bootstrap pro závislá data a detekce změn

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

praskova@karlin.mff.cuni.cz

Je známo, že pro závislá pozorování nelze přímo aplikovat klasický bootstrap. V tomto příspěvku budou uvedeny některé varianty metody bootstrap, které zohledňují závislostní strukturu dat. Pomocí těchto metod potom budou aproximovány kritické hodnoty testů pro detekci změn v parametrech lineárních modelů, jejichž chybové členy nejsou nezávislé. Přednosti a nedostatky jednotlivých metod budou zkoumány numericky.

Julie Rendlová

Analýza kategoriálních dat – problém vícenásobné volby v odpovědi

PřF UPOL, KMAAM, 17. listopadu 12, 771 46 Olomouc

julcika@seznam.cz

Příspěvek se zabývá analýzou kategoriálních dat v případě možnosti vícenásobné volby v odpovědi. Jeho cílem je seznámení s potřebným teoretickým zázemím a ukázka praktické aplikace zpracované v softwarovém prostředí statistického programu R, konkrétně s využitím balíčku MRCV.

Vysvětlující i vysvětlované proměnné se souhrnně označují zkratkou MRCV (multiple response categorical variables) a mohou nabývat libovolně mnoha kategorií nazývaných items, což jsou binární proměnné s hodnotami 0 a 1. Mezi items každé MRCV předpokládáme existenci vnitřních závislostí. S daty pracujeme v podobě item-response tabulek, což jsou kontingenční tabulky umožňující testování simultánní párové marginální nezávislosti. Teorie pro analýzu uvedené problematiky vychází z logaritmicko-lineárních modelů s použitím modifikované Pearsonovy statistiky a standardizovaných reziduí, Rao-Scottových korekcí a bootstrapových metod.

V praktické části příspěvku aplikujeme metody na typický příklad studovaného typu dat, jímž jsou dotazníky s možností výběru všech relevantních odpovědí na zadanou otázku. Protože při práci s takovými dotazníky neustále převládá užívání klasických logaritmicko-lineárních modelů, pokusíme se také vysvětlit, proč je tento přístup v daném případě nevhodný.

Samuel Rosa, Radoslav Harman

Optimal designs for dose-escalation studies

FMFI UK, KAMŠ, Mlynská dolina 6284, 842 48 Bratislava 4

samuel.rosa@fmph.uniba.sk

Consider an experiment, where a new drug is tested for the first time on human subjects. Such experiments are often performed as dose-escalation studies, where a set of increasing doses is pre-selected, individuals are grouped into cohorts and a dose is given to subjects in a cohort only if the preceding dose was already given to previous cohorts. If an adverse effect is observed, the experiment stops and thus no subjects are exposed to higher doses. We consider the model for dose-escalation studies formulated by [1]. Specifically, we assume that the response is affected both by the dose or placebo effects as well as by the cohort effects. We provide optimal approximate designs for selected optimality criteria (E -, MV - and LV -optimality) for estimating the effects of drug doses compared with placebo. In particular, we obtain the optimality of Senn designs and extended Senn designs with respect to multiple criteria.

Literature

- [1] Bailey, R. A. (2009): Designs for dose-escalation trials with quantitative responses, *Statistics in Medicine* **28**, pp. 3721–3738

Acknowledgement: The research was supported by the UK/214/2016 grant of the Comenius University in Bratislava.

Šárka Rusá, Arnošt Komárek

Bayesovská analýza tříúrovňového modelu mediace s ordinální odezvou

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

rusa@karlin.mff.cuni.cz

Tato statistická analýza zkoumá vliv personálního zajištění zdravotních sester (měřen jako poměr pacientů na jednu sestru) a pracovního prostředí zdravotních sester na spokojenost pacientů se zdravotní péčí v nemocnicích a možnou závislost tohoto efektu na vzdělání personálu. Analýza dat z rozsáhlého evropského výzkumu RN4CAST, jehož víceúrovňovou strukturu jsme brali v úvahu, je založena na bayesovském tříúrovňovém modelu mediace. Jako mediátor v tomto modelu figuruje míra toho, kolik zdravotnické péče sestra nestihla vykonat. Předpokládáme, že ordinální odezva pochází z nepozorovaného normálního rozdělení s neznámými mezemi pro jednotlivé hodnoty.

Iveta Selingerová, Stanislav Katina, Ivana Horová

Jádrové odhady jako alternativa (semi)parametrických modelů v analýze přežití

PřF MU, ÚMS, Kotlářská 2, 611 37 Brno

xselinge@math.muni.cz

Riziková funkce patří k velmi užitečným nástrojům při zpracování dat v analýze přežití. Vyjadřuje okamžitou pravděpodobnost výskytu události v následujícím časovém okamžiku. V praxi toto riziko může být ovlivněno dalšími charakteristikami, jako je pohlaví, věk, hladina nádorových markerů apod. Pro modelování této závislosti se používají parametrické či semiparametrické metody. Omezením těchto metod je však požadavek na splnění určitých předpokladů, jako je rozdělení času přežití, proporcionalita rizik či exponenciální závislost na kovariátě. V tomto ohledu jsou neparametrické metody více flexibilní. Vzhledem k vlastnostem rizikové funkce jsou pro její odhad nejvhodnější vyhlazovací techniky, mezi něž patří jádrové odhady. Cílem příspěvku je představit různé typy jádrových odhadů rizikové funkce a zhodnotit výhody a nevýhody jednotlivých přístupů za využití simulační studie. V rámci studie jsme se zaměřili na různé tvary rizikové funkce a na vliv zastoupení cenzorovaných pozorování.

Ondřej Sokol, Elena Kuchina

Složitost výpočtu horní meze rozptylu nad náhodnými daty

FIS VŠE, KE, Nám. W. Churchilla 4, 130 67 Praha 3

ondrej.sokol@vse.cz; elena.kuchina@vse.cz

Zabýváme se problematikou výpočtu horní meze výběrového rozptylu v případě, kdy nejsou k dispozici přesná data, ale pouze intervaly, ve kterých tato přesná data s jistotou leží. Obecně je nalezení horní meze výběrového rozptylu ze znalosti pouze intervalových dat NP-obtížná úloha, ale při splnění určitých podmínek kladených na vstupní data lze použít některý z efektivních algoritmů. V této práci je konkrétně zkoumán Fersonův algoritmus pro výpočet horní meze rozptylu intervalových dat z pohledu průměrné výpočetní složitosti při náhodně generovaných datech z běžných rozdělení. Algoritmus pracuje obecně v exponenciálním čase, kde exponenciální část

složitosti závisí na maximálním počtu zúžených intervalů, které mají společný alespoň jeden bod (zúženým intervalem je zde myšlen interval, jehož poloměr je vydělen celkovým počtem intervalů v množině dat). Na základě simulačních experimentů je zformulována hypotéza, že za určitých podmínek střední hodnota maximálního počtu překrytí zúžených intervalů roste logaritmicky s počtem intervalů. Složitost výpočtu horní meze rozptylu nad náhodnými daty je pak tedy polynomiální.

Poděkování: Příspěvek vznikl s podporou projektu IGA F4/63/2016 Interní grantové agentury VŠE v Praze.

Gábor Szűcs

Rekurentné triedy diskretných rozdelení pravdepodobnosti a odhadovanie ich parametrov

FMFI UK, KAMŠ, Mlynská dolina 6284, 842 48 Bratislava 4

Gabor.Szucs@fmph.uniba.sk

Tento príspevok sa zaoberá s rekurentnými triedami diskretných rozdelení pravdepodobnosti. Zameriava sa predovšetkým na takzvanú \mathcal{R}_k -triedu rozdelení a jej špeciálne podtriedy, pričom uvádza aj ich základné vlastnosti. Hlavným cieľom príspevku je predstaviť možnosti odhadovania parametrov jednotlivých rekurentných podtried rozdelení pravdepodobnosti na základe dátového vektora. Keďže sa jedná o atypicky definované rozdelenia, odhadovanie parametrov rekurentných podtried je obvykle založené na numerických metódach a napríklad na numerickej minimalizácii testovej štatistiky Kolmogorovovho-Smirnovovho testu, Cramérovho-von Misesovho testu, prípadne inej funkcie odhadovaných parametrov. V praktickej časti príspevku sú uvedené programové implementácie a porovnania spomínaných metód odhadovania parametrov v rámci štatistického softvéru *R*.

Pod'akovanie: Táto práca bola podporená vedeckým grantom VEGA 2/0047/15.

Daniel Ševčovič

Riccati transformation method for solving Hamilton-Jacobi-Bellman equation

FMFI UK, Mlynská dolina 6284, 842 48 Bratislava 4 Comenius University Bratislava

sevcovic@fmph.uniba.sk

In this talk we present recent results on application of the Riccati transformation for solving the evolutionary Hamilton-Jacobi-Bellman equation arising from the stochastic dynamic optimal allocation problem.

It turns out that the fully nonlinear Hamilton-Jacobi-Bellman equation governing evolution of the value function can be transformed into a quasi-linear parabolic equation. Its diffusion function is obtained as a value function of certain parametric convex optimization problem. We will show that point of discontinuity of this value function can be identified with transitions and changes of the optimal portfolio composition.

We prove existence of classical solutions to the HJB equation. A numerical solution is then constructed by means of an implicit iterative finite volume numerical approximation scheme. As an application we present results of computing optimal strategies for a portfolio investment problem for German DAX stock index. This is a joint work with S. Kilianova.

Tereza Šimková

Multivariate L-moment homogeneity test for spatially correlated data

PrF TUL, KAP, Studentská 1402/2, 461 17 Liberec 1

tereza.simkova@tul.cz

Identification of homogeneous regions is a key task in regional frequency analysis to obtain adequate estimates of a given event. Recently Chebana and Ouarda (2007) have extended the univariate L-moment homogeneity test of Hosking and Wallis (1997) to the multivariate case. However, the proposed multivariate L-moment homogeneity test assumes intersite independence, although examples from practice demonstrate that intersite correlation may be expected for some kinds of data. Hence, the testing procedure of Chebana and Ouarda (2007) needs to be generalized: D-vine copulas are utilised to model intersite dependence when generating synthetic regions to overcome the problem of presence of cross-correlation between stations. Performed Monte Carlo simulations illustrates how intersite dependence impacts the multivariate L-moment homogeneity test by significant reducing the value of the heterogeneity measure. The results of simulations demonstrate the superiority of the proposed modification over the original procedure of Chebana and Ouarda (2007) since it improves the heterogeneity detection and avoids of misspecification of a studied region.

Literature

- [1] F. Chebana and T. B. M. J. Ouarda, "Multivariate L-moment homogeneity test," *Water Resources Research*, vol. 43, W08406, doi:10.1029/2006WR005639, 2007.
- [2] J. R. M. Hosking and J. R. Wallis, *Regional Frequency Analysis: An Approach Based on L-Moments*, Cambridge University Press, Cambridge, UK, 1997.

Zdeněk Šulc

Metodologie hodnocení měr podobnosti pro kategoriální data na velkém množství datových souborů

FIS VŠE, KE, Nám. W. Churchilla 4, 130 67 Praha 3

zdenek.sulc@vse.cz

Součástí většiny článků zabývajících se hodnocením měr podobnosti bývá hodnocení jejich výsledků na vybraných reálných datových souborech, které obvykle pocházejí z dobře známých repositářů, a jsou tak srovnatelné s výsledky jiných výzkumníků. Někdy je ale potřeba prozkoumat, ve kterých konkrétních situacích daná míra podobnosti podává dobré výsledky a kdy nikoliv. V takových případech je vhodné využít analýzu založenou na generovaných souborech, která umožňuje tvorbu souborů s požadovanými vlastnostmi. Vzhledem k tomu, že je obvykle zapotřebí velkého množství generovaných souborů, vyvstává zde otázka, jak zpracovat takové množství souborů, aby bylo dosaženo jednoduše interpretovatelných výsledků. Proto tento příspěvek prezentuje metodologii hodnocení měr podobnosti pro kategoriální data, která umožňuje porovnání kvality shlukování založené na interních hodnotících kritériích na velkém množství datových souborů. Tato metodologie může být používána buď samostatně s předem připravenými datovými soubory, nebo společně s funkcí generující kategoriální data. Oba typické způsoby použití jsou v příspěvku demonstrovány.

Renáta Talská, Karel Hron, Jitka Machalová, Eva Fišerová

Kompoziční regrese s funkcí závisle proměnnou

PrF UPOL, KMAAM, 17. listopadu 12, 771 46 Olomouc

talskarenata@seznam.cz

Regresní analýza je hojně používaný statistický nástroj sloužící k modelování vztahu mezi závisle proměnnou a množinou vysvětlujících proměnných (prediktorů). Aplikace funkcionální regrese je nezbytná v případě, kdy alespoň jedna ze zmíněných proměnných má funkcionální charakter. Přehled možných přístupů ke konstrukci regresních modelů s funkcí závisle proměnnou a skalárními prediktory byl představen v [1]. Zajímavou se však stává situace, kdy závisle proměnná je hustotou rozdělení pravděpodobností, neboť prostor L^2 (integrovatelných funkcí s druhou mocninou), ve kterém je typicky funkcionální regrese prováděna, nebere v úvahu relativní charakter hustot. Cílem tohoto příspěvku je představit funkcionální regresní model s distribuční závisle proměnnou s využitím metodiky Bayesových prostorů, tj. přístupu, který respektuje geometrické vlastnosti distribučních dat [2, 3]. Speciálně, protože hustoty představují funkcionální data nesoucí pouze relativní informaci, podmínku jednotkového integrálu lze vnímat jen jako jednu z ekvivalentních reprezentací hustot (invariance na změnu měřítka). Aby bylo možné použít metody funkcionální regrese, které byly navrženy pro funkcionální data z prostoru L^2 , zejména pak ty, které jsou založené na B-splajnové reprezentaci, uchylujeme se k zobrazení Bayesova prostoru do L^2 prostřednictvím tzv. centrované log-podílové (log-ratio) transformace. Prezentované teoretické poznatky budou ilustrovány na reálných datech.

Literatura

- [1] J. Ramsay, B.W. Silverman (2005). *Functional Data Analysis*. Springer, Heidelberg.
- [2] K. Hron, A. Menafoglio, M. Templ, K. Hružová, P. Filzmoser (2016). Simplicial principal component analysis for density functions in Bayes spaces. *Computational Statistics and Data Analysis*, 94, 330–350.
- [3] K. G. van den Boogaart, J.J. Egozcue, V. Pawlowsky-Glahn (2014). Bayes Hilbert spaces. *Australian & New Zealand Journal of Statistics*, 56(2), 171–194.

Marie Turčičová

Modelování kovariance v asimilaci dat odhadem prostorové struktury gaussovského markovského náhodného pole

MFF UK, KPMS, Sokolovská 83, 186 75 Praha 8; ÚI AV ČR, Pod Vodárenskou věží 271/2, 18207 Praha 8

turcic@karlin.mff.cuni.cz

Chceme-li odhadnout kovarianční matici na základě náhodného výběru, jehož rozsah je velmi malý v porovnání s dimenzí jednotlivých členů, pak výběrová kovarianční matice není dobrým odhadem, neboť má velmi malou hodnotu a může obsahovat tzv. rušivé korelace. Tento problém nastává například při aplikaci filtračních algoritmů v předpovědních modelech počasí, kde je odhad kovarianční matice klíčovým prvkem kvalitní předpovědi, avšak rozdíl mezi dimenzí náhodných vektorů a rozsahem výběru je v řádu tisíců. Vylepšení odhadu kovarianční matice je možné po zavedení dodatečných předpokladů, které však nesmějí být pro strukturu náhodných vektorů příliš omezující. V příspěvku prezentujeme odhad kovarianční matice založený na identifikaci parametrů prostorové struktury gaussovského markovského pole z náhodného vzorku metodou maximální věrohodnosti. Kovarianční matice se nevytváří přímo, ale výpočty s kovariancí tohoto pole mohou být realizovány řešením nehomogenní stochastické difúzní rovnice metodou konečných prvků. Výsledné procedury pro násobení matice a vektoru mohou být použity v asimilaci dat přímo, nebo generováním většího počtu nových náhodných prvků s následným výpočtem vylepšené výběrové kovariance.

Kristýna Vaňkátová, Eva Fišerová

Shluková analýza ve směsích regresních modelů

PřF UPOL, KMAAM, 17. listopadu 12, 771 46 Olomouc

kristyna.vankatova@upol.cz; eva.fiserova@upol.cz

Směsi lineárních regresních modelů jsou pokročilým nástrojem regresní analýzy, který je schopen pracovat s heterogenními daty a odhadovat regresní parametry s využitím podmíněného rozdělení pravděpodobnosti vysvětlované proměnné. Toto rozdělení je dáno jako vážený součet přes všechny komponenty obsažené v dané směsi a obsahuje všechny neznámé parametry, které lze odhadnout pomocí metody maximální věrohodnosti. Vzhledem ke složitosti takového optimalizačního problému byl vyvinut EM algoritmus umožňující výpočet maximálně věrohodných odhadů z neúplných dat. Přesnost odhadů a vymezení jednotlivých komponent se dá v některých případech značně vylepšit zahrnutím doprovodné proměnné do modelu. Aplikace směsí regresních modelů lze nalézt v mnoha vědních odvětvích, včetně biologie, genetiky, medicíny či ekonomie, v jejímž rámci je možné pozorovat například vztah mezi roční starobní penzí poskytovanou státem a příjmem lidí starších 65 let.

Jakub Večeřa

Estimation of parameters in a planar segment process model with a biological application

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

vecera@karlin.mff.cuni.cz

In many applications systems of randomly dispersed segments in the plane or space are investigated. In biology such systems occur e. g. when using fluorescent imaging of actin stress fibres in human mesenchymal stem cells from bone marrow. In materials research the microstructure of fibre-reinforced composites contain segments of small thickness. Typically the objects are not distributed purely randomly, which would correspond to the mathematical model called Poisson segment process. The aim is therefore to build stochastic models which involve some kind of interactions, which is a broad notion. Basically, such interactions can be attractive or repulsive, but in practical problems these can combine in various scales and also interactions with external deterministic components is frequent.

Ondřej Vencálek

Zobecněné lineární modely nebo analýza kompozičních dat? Podobnosti a rozdílnosti

PřF UPOL, KMAAM, 17. listopadu 12, 771 46 Olomouc

ondrej.vencalek@upol.cz

Naše práce byla motivována praktickým problémem analýzy datového souboru obsahujícího údaje o počtech cyklistů hospitalizovaných po dopravní nehodě. Otázkou bylo, zda a jak se během 11 let, za něž jsou údaje k dispozici, změnilo zastoupení tří věkových skupin mezi takto hospitalizovanými. Tradiční přístup k analýze takového typu dat je založen na použití metodologie zobecněných lineárních modelů pro ordinální závisle proměnnou. Nově se ale prosazuje pohled na data tohoto typu jako na tzv. kompoziční data. Tento přístup je zaměřen na analýzu informací o podílu jednotlivých částí na celku.

V příspěvku ukážeme nejprve podobnosti obou přístupů. Jedním ze zobecněných lineárních modelů je tzv. *adjacent-categories logit model*, v němž jsou modelovány logity pravděpodobností po sobě jdoucích kategorií

$$\ln \frac{\pi_1}{\pi_2}, \ln \frac{\pi_2}{\pi_3}.$$

Analogické výrazy (v nichž jsou pravděpodobnosti nahrazeny relativními četnostmi) mohou být použity i coby logratio souřadnice v rámci analýzy kompozičních dat. Zde jsou však preferovány poněkud odlišné, ortonormální souřadnice. Ty odpovídají výrazům

$$\ln \frac{\pi_1}{\pi_2}, \ln \frac{\sqrt{\pi_1 \pi_2}}{\pi_3}.$$

Závislost těchto souřadnic na čase (či jiné vysvětlující proměnné) je modelována pomocí lineárního regresního modelu. Podstatné rozdílnosti v obou přístupech pomůže objasnit vyšetření asymptotických vlastností logratio souřadnic. . .

Mojmír Vinkler, Stanislav Katina, Miroslav Smíšek

Effect of denoising on brain atrophy measurements based on MRI for Alzheimer's disease

MV, SK: PrF MU, ÚMS, Kotlářská 2, 611 37 Brno

MS: AXON Neuroscience SE

mojmir.vinkler@gmail.com

Alzheimer's disease (AD) is a neurocognitive disorder with various level of atrophy of cerebral cortical regions and subcortical structures - mainly hippocampus. Volumetric measurements of hippocampus and its atrophy rate are promising tools in the determination of effect of disease-modifying treatment. Automatic segmentation of various cerebral structures visualized by MR and measurement of their volume is now freely available for clinical research of the natural course of the disease and effect of the treatment. However, due to inherent noise, variance of these measurement is high, making difficult reaching statistically valid conclusions. Using 140 MRI scans (28 patients, 5 visits) from randomized, placebo-controlled, parallel group, double-blinded, multi-centre Phase I clinical study we found that denoising with state of the art method prior to running FreeSurfer automatic segmentation reduced measurement error from 6.79% to 3.54% without introducing processing bias. Furthermore, additional temporal information reduced error to 2.50%. Lower error significantly reduced sample size required to detect differences in hippocampus atrophy between control (placebo) and treated group and also between left and right hippocampus. Although our results are not significant due to short study length and small sample size, general trend looks promising with respect to design, i.e. power analysis and sample size calculation, of further Phase II and III clinical studies.

Jan Ámos Víšek

Are the bad leverage points the most difficult problem for estimating the underlying regression model?

FSV UK, IES, Opletalova 26, 110 01 Praha 1

visek@fsv.cuni.cz

A series of unsuccessful proposals of robust estimators hopefully with 50% breakdown point led to Pyrrhic victory by Andrew Siegel's *repeated median*, see [17]. Nevertheless, feasible versions of 50% breakdown point estimators - the *least median of squares* (LMS) (see [14]) and the *least trimmed squares* (see [9]¹) - were announced nearly immediately after it. They fulfilled the desire but the discontinuity of objective function and the presence of order statistics of the squared residuals in their definitions have made it difficult to study the properties of estimators in question. The proof of consistency of the former still exists only for special case and for the fully general proof of consistency of the latter we waited more than 20 years (and the proof is awfully involved, see [22]). On the other hand, efficient algorithms² for computing both of them arrived rather soon.

The proposal of *S-estimator* (see [16]) removed both these snags simultaneously preserving the high breakdown point and delivering simultaneously the proof of consistency, based on the results of [13]. A tax we had to pay for it was a restriction of the range of objective functions, but the restriction was quite acceptable. It required the objective function to be symmetric around zero and constant starting with some $c > 0$.

The algorithms for computing *S-estimators* were implemented also nearly immediately but - due to the fact that they represented rather intensive computing - they are still studied (see e.g. [5], [6] and [19]). The extraordinary virtue of all these estimators was their "innate" scale- and regression-equivariance in contrast to

¹This reference on LTS is usually given although Peter Rousseeuw announced LTS only a few month after LMS. On the other hand, the first remark on possibility to compute something like LTS is in [8].

²The algorithms for LMS (see [4]) and LTS (see [1] or [23]) are efficient in a non-statistical sense, i.e. that they give hopefully tight approximations to the exact solutions of the corresponding extremal problems. The hope for it is supported by following: The estimate of underlying regression model for the group of datasets which has become benchmarks in robust regression, see [15], are such that the h -th order statistic of squared residuals in *LMS*-estimate is smaller than the h -th order statistic of squared residuals in *LTS*-estimate and vice versa, the sum of the first h -th order statistics of squared residuals in *LTS*-estimate is smaller than the sum of the first h -th order statistics of squared residuals in *LMS*-estimate. And it holds even for the case when we are able to give for LTS the exact solution, see [10] and [20].

M -estimators which require studentization, moreover the studentization by rather special estimator of standard deviation of error term³. The estimator of scale has to be scale-equivariant and regression-invariant, see [3] and [11].

Shocking results by Thomas Hettmasperger and Simon Sheather (see [10], although they later proved to be wrong due to the bad algorithm they used for computing LMS, see [20]) revealed that the requirement on high breakdown point and/or zero-one objective function inevitably results in the high sensitivity to a small shift of “in-liers”, the drawback which cannot be removed without releasing the requirement on the extremely high breakdown point and zero-one objective function.

That was the reason why an attempt of coping with this problem by offering a possibility of accommodating the level of robustness of the estimator to the level of contamination - the *least weighted squares* (LWS) - appeared (see [21]). The accommodation was enabled by appropriate adjustment of weights. In fact, the high speed of modern computational means allowed to select the weights-generating-function just tailored to the level and even to the character of contamination by a *forward search*, see [2]. It is plausible from the applications-point-of-view. The utilization of a generalized version of Glivenko theorem [7] (or Kolmogorov-Smirnov results, [12] or [18]) about the convergence of empirical distribution functions to the underlying one - generalized for the regression framework, see [24] - then simplified (in the sense of employment of much simpler and really applicable tools than the theory of empirical processes offers) the proofs of consistency, \sqrt{n} -consistency, etc.

Nevertheless, a disadvantage of LWS, mostly from the theoretical point of view, was the objective function - the only one, the quadratic function. Allowing for a general objecting function (even for an unbounded one) yielded the *S-weighted estimators*, see [25]. They inherited plausible properties of *S-estimators* as well as of *LWS*⁴, opening a chance to utilize a wide range of objective functions and simultaneously offering - by an appropriate selection of weight function - to adjust the estimator to the level and to the character of contamination.

The contribution summarizes the conditions for the consistency, \sqrt{n} -consistency (even under heteroscedasticity of error terms), asymptotic representation of *SW*-estimators (see [26]) and discusses the possibility of a “new” algorithm for computing them. The converted commas indicate that the trick for a plausible way for computing the estimator was in fact already employed in nineties for computing the M -estimators and is due to Jaromír Antoch, see [1]. A small collection of patterns of results of numerical studies of their behavior for moderate sample sizes will conclude the contribution. This is the most interesting part of the contribution because the results indicate that the widely spread idea that the most dangerous contamination for estimating regression model is represented by the leverage points need not be - sometimes - valid.

Literatura

- [1] Antoch, J., J. Á. Víšek (1991): Robust estimation in linear models and its computational aspects. *Contributions to Statistics: Computational Aspects of Model Choice*, Springer Verlag, (1992), ed. J. Antoch, 39 - 104.
- [2] Atkinson, A. C., M. Riani, A. Cerioli (2004): *Exploring Multivariate Data with the Forward Search*. Springer Series in Statistics 2004, 31-88.
- [3] Bickel, P. J. (1975): One-step Huber estimates in the linear model. *JASA* 70, 428-433.
- [4] Boček, P., P. Lachout (1993): Linear programming approach to *LMS*-estimation. *Memorial volume of Comput. Statist. & Data Analysis 19(1995)*, 129 - 134.
- [5] Campbell, N. A., Lopuhaa, H. P., Rousseeuw, P. J. (1998): On calculation of a robust *S*-estimator of a covariance matrix. *Statistics in medicine*, 17, 2685 - 2695.
- [6] Desborges, R., Verardi, V. (2012): A robust instrumental-variable estimator. *The Stata Journal (2012)* 12, 169 -181.
- [7] Glivenko, V. I. (1933): Sulla determinazione empirica delle leggi di probabilita. *Giorn. Ist.Ital. Attuari* 4, 92.
- [8] Hampel, F. R. (1968): *Contributions to the theory of robust estimation*. Ph. D. thesis. University of California, Berkeley.
- [9] Hampel, F. R., E. M. Ronchetti, P. J. Rousseeuw, W. A. Stahel (1986): *Robust Statistics – The Approach Based on Influence Functions*. New York: J.Wiley & Son.
- [10] Hettmasperger, T. P., S. J. Sheather (1992): A Cautionary Note on the Method of Least Median Squares. *The American Statistician* 46, 79-83.
- [11] Jurečková, J., P. K. Sen (1984): On adaptive scale-equivariant M -estimators in linear models. *Statistics and Decisions* 2, Suppl. Issue No. 1.
- [12] Kolmogorov, A. (1950): *Foundations of Probability*. (English translation) Chelsea Publishing Co., New York, 1950.
- [13] Maronna, R. A., V. J. Yohai (1981): Asymptotic behaviour of general M -estimates for regression and scale with random carriers. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* 58, 7-20.

³Of course, it implicitly says that M -estimators assume the homoscedastic error terms which - especially for applications in social sciences - represents rather serious restriction.

⁴It is rather simple to show that *S-estimators* are not special case of *LWS* and vice versa.

- [14] Rousseeuw, P.J. (1984): Least median of square regression. *JASA* 79, 871-880.
- [15] Rousseeuw, P. J., A. M. Leroy (1987): *Robust Regression and Outlier Detection*. New York: J.Wiley & Sons.
- [16] Rousseeuw, P. J., V. Yohai (1984): Robust regression by means of S -estimators. In: *Robust and Nonlinear Time Series Analysis*. eds. J. Franke, W. Härdle and R. D. Martin, *Lecture Notes in Statistics No. 26 Springer Verlag, New York, 256-272*.
- [17] Siegel, A. F. (1982): Robust regression using repeated medians. *Biometrika*, 69, 242 - 244.
- [18] Smirnov, N. (1939): On the estimation of discrepancy between empirical curves of distribution for two independent samples. *Bull. Math. Univ. Moscow* 2, 3 - 14.
- [19] Verardi, V., McCathie, A. (2012): The S -estimator of multivariate location and scatter in Stata. *The Stata Journal* (2012) 12, 299 - 307.
- [20] Víšek, J. Á. (1994): A cautionary note on the method of Least Median of Squares reconsidered, *Transactions of the Twelfth Prague Conference on Information Theory, Statistical Decision Functions and Random Processes*, Lachout, P. (ed), *Academy of Sciences of the Czech Republic, Prague, 1994, 254 - 259*.
- [21] Víšek, J. Á. (2000): Regression with high breakdown point. *Robust 2000* (eds. J. Antoch & G. Dohnal, *Union of Czech Mathematicians and Physicists*), Prague: matfyzpress, 324 - 356.
- [22] Víšek, J. Á. (2006): The least trimmed squares. Part I - Consistency. Part II - \sqrt{n} -consistency. Part III - Asymptotic normality and Bahadur representation. *Kybernetika* 42, 1 - 36, 181 - 202, 203 - 224.
- [23] Víšek, J. Á. (2006): Instrumental Weighted Variables - algorithm. *Proceedings of the COMPSTAT 2006*, eds. A. Rizzi & M. Vichi, *Physica-Verlag, 777-786*.
- [24] Víšek, J. Á. (2011): Empirical distribution function under heteroscedasticity. *Statistics* 45, 497-508.
- [25] Víšek, J. Á. (2015): S -weighted estimators. *Proceedings of the 16th Conference on the Applied Stochastic Models, Data Analysis and Demographics 2015*, 1031 - 1042.
- [26] Víšek, J. Á. (2016): Asymptotics of S -weighted estimators. Submitted to *Contributions to Theoretical and Applied Statistics In honor of Corrado Gini*.

Acknowledgement: Research was supported by the Czech Science Foundation project No. 13-01930S Robust methods for nonstandard situations, their diagnostics and implementations.

Petr Volf

O analýze konkurujících si rizik s aplikací na čas prvního gólu v fotbalovém utkání

ÚTIA AV ČR, Pod vodárenskou věží 4, 18208 Praha 8

volf@utia.cas.cz

V příspěvku nejprve zopakují základní poznatky o problému konkurenčních rizik. Ten se vyskytuje v statistické analýze přežití a zobecňuje vlastně schema náhodného cenzorování zprava v tom smyslu, že konkurenční veličiny (zde tedy zároveň běžící náhodné časy do "konkurenčních" událostí, tj. takových, že z nich může nastat jen jedna) mohou být vzájemně závislé. V medicínských problémech se většinou analyzuje incidence (tj. skutečný výskyt) takovýchto událostí, ale zkoumání vzájemné souvislosti je zajímavé jednak prakticky, a jednak i z hlediska modelů a analýzy. Nyní se na modelování závislosti náhodných veličin používají modely kopulí, tento trend pronikl i sem. Díky neúplně pozorovaným datům se ovšem může stát, že kompletní model není identifikovatelný. Další otázkou je výběr kopuly, vhodnost lze ověřit dodatečným testováním shody modelu s daty. I těmto otázkám se budu věnovat. Hlavním cílem je pak zkusit model konkurenčních rizik uplatnit na "konkurující si" 2 náhodné časy (2 týmů) do vstřelení první branky ve fotbalovém utkání. Proto také připomenou základní přístupy k modelování skóre (i jeho vývoje) zápasu a rozšířím je právě o uvažování závislosti intenzit skórování obou týmů. Praktická analýza bude provedena na datech z jednoho ročníku české Synot ligy.

Viktor Witkovský

Vybrané metody a aplikácie štatistickej inferencie založené na numerickej inverzii charakteristickej funkcie

Ústav merania SAV, Dúbravská cesta 9, 841 04 Bratislava

witkovsky@savba.sk

Cieľom príspevku je prezentovať a ilustrovať vybrané metódy, algoritmy a ich aplikácie pre štatistickú inferenciu založenú na kombinovaní a invertovaní charakteristických funkcií (CF). Tradične sú charakteristické funkcie využívané ako jeden zo základných teoretických nástrojov pravdepodobnosti a matematickej štatistiky. Práca s charakteristickými funkciami je častokrát výhodnejšia a jednoduchšia oproti postupom založeným na priamej manipulácii s distribučnými funkciami (CDF) resp. s ich hustotami rozdelenia (PDF). Menej frekventované je však priame využitie charakteristických funkcií vo výpočtovej resp. aplikovanej štatistike. Jedným z dôvodov je absencia spoľahlivých výpočtových nástrojov pre prácu s charakteristickými funkciami v štandardných softvérových

balíčkoch (napr. R, SAS, MATLAB). Vo všeobecnosti, problematika numerického invertovania CF naráža na komplikácie spojené s problémom integrovania oscilujúcich funkcií. Pokiaľ však zostaneme na pôde aplikovanej štatistiky, častokrát už veľmi jednoduché numerické nástroje poskytujú dostatočne presné (aproximatívne) výsledky.

V príspevku uvedieme jednoduché metódy a algoritmy pre numerické invertovanie charakteristických funkcií, založené na Gil-Pelaezovej metóde invertovania CF a na použití algoritmu pre rýchlu Fourierovú transformáciu (FFT). Aplikovateľnosť týchto metód ilustrujeme na niekoľkých príkladoch parametrickej ako aj neparametrickej štatistickej inferencie (založenej na empirických charakteristických funkciách).

PodĎakovanie: Táto práca bola podporená Agentúrou na podporu výskumu a vývoja APVV-15-0295.

Xeniya Yermolenko

Non-unbiased two-sample nonparametric tests: Numerical example

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

yermolenko@karlin.mff.cuni.cz

Many tests on vector or scalar parameters against two-sided alternatives are generally not finite-sample unbiased. They are unbiased only for symmetric distributions or under similar conditions. This was already noticed by [1], [4] and generally analyzed by [2], [3] and later by many others. While in univariate models the tests are unbiased against one-sample alternatives, such alternatives are not clearly characterized in the multivariate models.

We shall numerically illustrate this important problem on the Wilcoxon test against two-sample alternative of shift in location, applied to a skew logistic distribution and unequal sample sizes.

Literature

- [1] Amrhein, P. (1995). An example of a two-sided Wilcoxon signed rank test which is not unbiased. *Ann. Inst. Statist. Math.* **47** 167–170.
- [2] Jurečková, J., & Kalina, J. (2012). Nonparametric multivariate rank tests and their unbiasedness. *Bernoulli* **18**, 229–251.
- [3] Jurečková, J. & Milhaud, X. (2003). Derivative in the mean of a density and statistical applications. *IMS Lecture Notes* **42**, 217–232.
- [4] Sugiura, N., Murakami, H., Lee, S.K. & Maeda, Y. (2006). Biased and unbiased two-sided Wilcoxon tests for equal sample sizes. *Ann. Inst. Statist. Math.* **58**, 93–100.

Markéta Zikmundová

Procesy interagujících úseček

VŠCHT, Ústav matematiky, Studentská 6, 160 00 Praha 6

marketa.zikmundova@vscht.cz

Uvažujme bodový proces interagujících úseček na omezené množině $S \subset \mathbb{R}^2$ daný hustotou

$$p(\mathbf{y}|\mathbf{x}) = c_{\mathbf{x}}^{-1} \exp(x_1 \cdot N(U_{\mathbf{y}}), x_2 \cdot L(U_{\mathbf{y}}), x_3 \cdot I(U_{\mathbf{y}})),$$

vzhledem k nějakému Poissonovu bodovému procesu úseček se součinnovou mírou intenzity. Symbol $N(U_{\mathbf{y}})$ značí počet průsečíků všech úseček ve sjednocení U konfigurace úseček \mathbf{y} . Obdobně L odpovídá celkové délce v U , I počtu izolovaných úseček a $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ je parametr rozdělení hustoty p . Tento příspěvek se zabývá odhadem parametru \mathbf{x} a odhadem vybraných charakteristik.