



## HORA MATKY BOŽÍ - KLÁŠTER KRÁLÍKY

30. ledna – 5. února 2010

ROBUST 2010 • 16. zimní škola JČMF a ČStS

# ROBUST 2010

## Sborník abstraktů

Antoch Jaromír	
<i>O testování shody ROC křivek</i> .....	7
Arendacká Barbora	
<i>Heteroskedastická jednofaktorová ANOVA</i>	
<i>intervalů pro varianční komponenty</i> .....	7
Bartošová Jitka a Forbelská Marie	
<i>Modelování rozdělení ročních příjmů u českých domácností</i> .....	8
Běláček Jaromír	
<i>Lateralita z pohledu ROC analýzy</i> .....	8
Branda Martin	
<i>Metody robustní statistiky ve stochastické optimalizaci</i> .....	9
Bubelíny Peter	
<i>Hotellingov test pro závislé data</i> .....	9
Cimermanová Katarína	
<i>Klasifikácia zašumených dát</i> .....	10
Dienstbier Jan	
<i>Tail modelling in linear models by quantile regression</i> .....	11
Dohnal Gejza	
<i>Modely sousledných událostí</i> .....	12
Došlá Šárka	
<i>Estimation of parameters of a clipped MA(1) process</i> .....	12
Dvořák Marek	
<i>On testing changes in parameters of an autoregressive model</i> ....	13
Fabián Zdeněk	
<i>Míry a váhy</i> .....	13
Filová Lenka	
<i>Minimal efficiency of designs in quadratic regression</i>	
<i>on q-dimensional cube</i> .....	14
Franc Jiří	
<i>Robustifikované instrumentální proměnné</i> .....	14
Frcalová Blažena	
<i>Časoprostorové bodové procesy</i> .....	15
Friesl Michal	
<i>Konzistence neparametrického bayesovského odhadu</i> .....	15

Grendár Marian	
<i>Empirická vierohodnosť</i> .....	15
Helisová Kateřina	
<i>Silová mozaika jako nástroj pro odhad parametrů v modelu</i>	
<i>náhodné množiny</i> .....	16
Hlávka Zdeněk	
<i>O neparametrickém odhadu polohy maxima</i> .....	17
Hlubinka Daniel, Kvitkovičová Andrea	
<i>Odhady a testy pro parametry lineárně unášeného</i>	
<i>Wienerova procesu</i> .....	18
Hornišová Klára	
<i>Neparametrická kalibrácia</i> .....	18
Hron Karel	
<i>Robustní metody pro kompoziční data</i> .....	19
Hušková Marie	
<i>Change point detection in trending regression</i> .....	20
Hykšová Magdalena	
<i>Filozofické interpretace pravděpodobnosti v díle</i>	
<i>T.G. Masaryka a K. Vorovky</i> .....	21
Chochola Ondřej	
<i>Sekvenční monitorování v kvantilové regresi</i> .....	22
Chvosteková Martina	
<i>Simultánne tolerančné intervaly v lineárnom regresnom modeli</i> ..	23
Janáček Jiří	
<i>Variance odhadů plochy a délky pomocí periodických mřížek</i> ....	24
Janková Janka	
<i>Intervaly spoľahlivosti pre spoločnú strednú hodnotu</i>	
<i>porovnanie dvoch metód</i> .....	25
Jarušková Daniela	
<i>Metody pro aproximaci horních kvantilů lineární kombinace</i>	
<i><math>\chi^2</math> rozdělení</i> .....	26
Jonáš Petr	
<i>Robustní odhad vícerozměrného modelu lineární regrese</i> .....	26
Jurczyk Tomáš	
<i>Vliv multikolinearity a odlehlých pozorování</i> .....	27
Jurečková Jana	
<i>Asymptotika versus konečně mnoho pozorování</i> .....	27
Juríček Jozef	
<i>Maximization of the information divergence from</i>	
<i>multinomial distributions</i> .....	28

Kalousová Anna	
<i>Joseph Bertrand</i> .....	28
Kaluža Jan	
<i>Maximální nerovnost pro stochastickou konvoluci řízenou martingalem</i> .....	29
Klaschka Jan	
<i>O výpočtu Blakerova konfidenčního intervalu</i> .....	29
Klein Daniel, Žežula Ivan	
<i>Orthogonal decompositions in growth curve models</i> .....	30
Komárek Arnošt	
<i>Sdružené modelování spojitých i diskrétních longitudinálních dat s exkurzí do diskriminační a shlukové analýzy</i> .....	30
Konár Ondřej	
<i>Detekce zvýšených ztrát v distribuční síti zemního plynu</i> .....	31
Kotík Lukáš	
<i>Lokální regresní hloubka</i> .....	31
Kraus Jan	
<i>Inference druhého řádu pro gaussovské náhodné křivky s aplikací na geometrii minikroužků DNA</i> .....	32
Kulich Michal	
<i>Analýza stratifikovaných dvoufázových studií s kalibrovanými a odhadnutými vahami</i> .....	32
Kvaszová Milena	
<i>Jak studenti rozumějí základním statistickým pojmům</i> .....	33
Lachout Petr	
<i>Aproximativní řešení a hodnota účelové funkce</i> .....	33
Lechnerová Radka, Lechner Tomáš	
<i>Aplikace bodových procesů při analýze veřejné správy v ČR</i> .....	34
Linka Aleš, Kula Jiří, Tunák Mároš	
<i>Detekce defektů plošných struktur</i> .....	34
Maciak Matúš	
<i>Bootstrapping of M-smoothers</i> .....	35
Madurkayová Barbora	
<i>Ratio type statistics for detection of changes in mean and the bootstrap method</i> .....	36
Marek Jaroslav, Kubáček Lubomír	
<i>O jednom geodetickém problému</i> .....	36
McLoone Jon	
<i>New statistical features of Mathematica</i> .....	36

Nagy Stanislav	
<i>Hĺoubka funkcionálnich dat</i> .....	37
Navrátil Radim	
<i>Chování pořadových testů v lineárním modelu za přítomnosti chyb měření</i> .....	38
Novák Petr	
<i>Testy dobré shody pro model zrychleného času v analýze přežití</i> ...	38
Pawlas Zbyněk	
<i>Odhad rozdělení dob mezi událostmi z krátkých časových oken</i> ...	38
Pešta Michal	
<i>Konzistentné a ekvivariantné odhadovanie v modeli so závislými chybami v premenných</i> .....	39
Petrásek Jakub	
<i>Modelling with jump processes and optimal control</i> .....	40
Picek Jan	
<i>Odhady parametrů v modelu s chybami měření</i> .....	40
Prášková Zuzana	
<i>Bootstrap v RCA modelech</i> .....	41
Sabolová Radka	
<i>Testy normality za přítomnosti rušivej regresie</i> .....	41
Schlesinger Pavel	
<i>Využití Gibbsova algoritmu ve shlukovacích úlohách komputační lingvistiky</i> .....	42
Shokirov Bobosharif, Klebanov Lev	
<i>Estimating the proportion of false hypotheses in multiple testing procedure</i> .....	42
Šedová Michaela, Kulich Michal	
<i>Dvoustupňové náhodné výběry ve výběrových šetřeních</i> .....	43
Šnupárková Jana	
<i>Stochastická bilineární rovnice s frakcionálním šumem v nekonečné dimenzi</i> .....	44
Staněk Jakub	
<i>Difúze v omezené oblasti s odrážející nebo pohlcující bariérou</i> ...	44
Timková Jana	
<i>On Bernstein-von Mises theorem and survival analysis</i> .....	44
Tuček Pavel a kol.	
<i>Design of measurement and statistical processing of magnetization analysis</i> .....	45
Vencálek Ondřej	
<i>Klasifikace na základě hloubky bodu</i> .....	45

Víšek Jan Ámos	
<i><math>\sqrt{n}</math>-Consistency of the least weighted squares     under heteroscedasticity of error terms</i>	46
Volf Petr	
<i>On models for progression of record values</i>	47
Wimmer Gejza Jr.	
<i>Konfidenčné oblasti pre regresné parametre v lineárnom     zmiešanom modeli pre longitudinálne dáta</i>	48
Witkovský Viktor	
<i>Využitie funkcie vierohodnosti na konštrukciu približných     a exaktných konfidenčných oblastí</i>	49
Zikmundová Markéta	
<i>Použití částicového filtru na odhad podmíněné intenzity     Hawkesova procesu</i>	50
Žambochová Marta	
<i>Shlukování v souborech s odlehlými objekty pomocí metod     k-průměrů</i>	50
Žežula Ivan	
<i>Logistic, multinomial, and ordinal regression</i>	50

**Antoch Jaromír****O testování shody ROC křivek**

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

jaromir.antoch@mff.cuni.cz

Předpokládejme, že je dána množina objektů, které jsou charakterizovány určitými měřitelnými vlastnostmi. Stojíme-li před problémem, jak dané objekty co nejlépe roztrdit do skupin, potřebujeme stanovit rozhodovací pravidlo (klasifikátor), které by dokázalo dobře využít znalost vlastností jednotlivých objektů a na základě této znalosti dospělo k jejich roztrdění.

Rozřazujeme-li objekty do předem známých tříd, jedná se o tzv. učení (se) s učitelem (supervised learning). V příspěvku se soustředíme na nalezení testů shody ROC křivek, které by umožnily vzájemné porovnání příslušných klasifikátorů, a prozkoumání vlastností nalezených testů. Motivaci pro řešení tohoto problému lze nalézt všude tam, kde je nutno porovnat více dostupných klasifikačních metod. Příkladem může být zavádění nových léčebných postupů v medicíně, počítačová lingvistika, apod.

**Arendacká Barbora****Heteroskedastická jednofaktorová ANOVA****intervaly pre variančné komponenty**

Ústav merania, Slovenská akadémia vied, Dúbravská cesta 9,

SK-841 04 Bratislava

barendacka@gmail.com

V príspevku sa budeme zaoberať intervalmi spoľahlivosti pre variančné komponenty v heteroskedastickom jednofaktorovom ANOVA modeli s náhodnými efektmi. Pôjde najmä o intervaly pre varianciu náhodného faktora. Zjednocujúci pohľad na už navrhnuté procedúry [1, 2] odhalí ďalšie možnosti ako voliť intervaly založené na metóde zovšeobecenej inferencie a tiež poukáže na určitú analógiu s homoskedastickým prípadom modelu. Vlastnosti vybraných intervalov ilustrujeme pomocou simulácií.

**Reference**

- [1] Li X. (2007) Comparison of confidence intervals on between group variance in unbalanced heteroscedastic one-way random models. *Commun. in Stat. Simulation and Computation* **36**, 381–390.
- [2] Wimmer G., Witkovský V. (2003) Between group variance component interval estimation for the unbalanced heteroscedastic one-way random effects model, *J. of Stat. Computation and Simulation* **73**, 333–346.

*Pod'akovanie:* Práca bola podporená grantom č. LPP-0388-09 poskytnutým Agentúrou na podporu výskumu a vývoja

## **Bartošová Jitka a Forbelská Marie**

### **Modelování rozdělení ročních příjmů u českých domácností**

KMI FM VŠE, Jarošovská 1117/II, CZ-377 01 Jindřichův Hradec

ÚMS PČF MU, Kotlářská 2, CZ-611 37 Brno

bartosov@fm.vse.cz, forbel@math.muni.cz

Příspěvek je věnován statistickým metodám, pomocí kterých lze modelovat rozdělení ročních příjmů českých domácností. Charakteristickým znakem tohoto rozdělení je mírné zešíkmení ve střední části s prudkým poklesem hustoty na obou stranách, s krátkým a masivním levým okrajem a dlouhým chvostem na pravé straně. V některých sociálních skupinách je toto rozdělení vícemodální, proto se přirozeně nabízí využít k modelování kromě známých rozdělení také konečné směsi, ať už se jedná o konečné směsi hustot či zobecněných lineárních modelů (GLM).

*Poděkování:* Příspěvek vznikl jako součást řešení grantového projektu GAČR 402/09/0515.

## **Běláček Jaromír, Tichý Jiří, Pecha Ondřej, Novák Marek**

### **Lateralita z pohledu ROC analýzy**

ÚBI 1. LF UK, BIO-STAT, Praha,

Neurologická klinika 1.LF UK Praha a VFN

jaromir.belacek@lf1.cuni.cz

Lateralitou rozumíme označení pro asymetrii ve větší fyziologické angažovanosti levé či pravé části lidského těla při různých aktivitách, jmenovitě u horních a dolních končetin, u mozkových hemisfér (řečové a symbolické funkce) nebo např. u párových smyslových orgánů (okulární či vestibulární dominance). Měření lateralit se tradičně provádí buď testováním vybraných položek prostřednictvím dotazníku (zda jsou úkony prováděny pravou či levou končetinou nebo oběma), monitorováním mozku prostřednictvím funkční magnetické rezonance (fMRI) či jinou specializovanou zobrazovací lékařskou technikou. Statistické vyhodnocení výsledků těchto měření lze odvíjet od jednoduchých měr (pravé či levé) lateralit a jejich komparací, testů chí-kvadrát nezávislosti v kontingenčních tabulkách 2x2 anebo 3x3 (včetně analýzy adjustovaných reziduí). Ve většině případů jde však o testování signifikantně



závislých indikátorů, kde běžné statistické testy (např. symetrie nebo prostřednictvím Spearmanových korelačních koeficientů) neumožňují diskriminovat či predikovat laterální položky (např. „ve vztahu k diagnóze rukosti /handedness/“). Za výše uvedeným účelem jsme začali experimentovat se standardní metodikou ROC analýzy, která přináší v laterálním kontextu zajímavé interpretace ukazatelů běžných u ROC (senzitivity, specificity, PPV a NPV) a specifickým způsobem kvantifikuje rozdíly mezi predikční schopností fenoménů měřených u praváků a u leváků, které jsme sice intuitivně „cítili“, ale nedokázali dosud výše jmenovanými metodickými prostředky formalizovat nebo transparentně vizualizovat. Přednáška bude podložena řadou ilustračních obrázků, které demonstrují výsledky dosažené při řešení projektu MSM 0021620816.

## Reference

- [1] Tichý J., Běláček J. (2009) Laterality in children: cerebellar dominance, handedness, footedness and hair whorl. *Activitas Nervosa Superior REDIVIVA. Act Nerv Super Rediviva* **51 (1-2)**, 2009, Slovak Academy of Sciences & St. Elisabeth University College of Health and Social Work, ISSN 1337-933X, 9-20.

## Branda Martin

### Metody robustní statistiky ve stochastické optimalizaci

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

branda@karlin.mff.cuni.cz

Příspěvek se zabývá možnostmi aplikovat metody a výsledky robustní statistiky v úlohách stochastického programování. Robustifikace úlohy probíhá skrze účelovou funkci, která závisí na pravděpodobnostním rozdělení a přípustném rozhodnutí. Jsou rozšířeny pojmy influenční funkce (influence function) a bod selhání (breakdown point) pro optimální hodnotu optimalizační úlohy. Je diskutován jejich výpočet a uvedeny příklady.

## Bubelíny Peter

### Hotellingov test pre závislé dáta

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

bubeliny@karlin.mff.cuni.cz

Práce je motivovaná analýzou génových množin, speciálně hľadáním odlišne expresovaných génových množin medzi dvomi skupinami. Génové expresie

sú silne závislé a ich  $\log_2$  transformácie majú približne normálne rozdelenie. Preto sa zdá byť rozumné použiť dvojjvýberový Hotellingov test pre tieto dáta. V tejto práci je vyšetované správanie tohto testu pre závislé dáta. Zdá sa, že tento test má niektoré neočakávané vlastnosti, ktoré ho robia odlišným od väčšiny testov používaných pre génové množiny. Ukázalo sa, že Hotellingov test nemusí mať najväčšiu silu, keď všetky marginálne rozdelenia sú odlišné. Pre silne závislé dáta je dosiahnutá jeho maximálna sila, keď je iba okolo polovica marginálnych rozdelení odlišných. Ďalšia zaujímavá vlastnosť je, že pre korelačný koeficient väčší ako 0.5 je tento test silnejší ak je iba jedno marginálne rozdelenie odlišné, ako keď sú všetky marginálne rozdelenia rovnako posunuté. Navyše s rastúcim korelačným koeficientom sila tohto testu rastie.

## Cimermanová Katarína

### Vplyv šumu na klasifikáciu dát do dvoch tried

Ústav merania, Slovenská akadémia vied, Dúbravská cesta 9,  
SK-841 04 Bratislava

katarina.cimermanova@gmail.com

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

Robustnú klasifikačnú metódu aplikujeme na analýzu vydychovaných plynov, kde sa budeme venovať klasifikácii dobrovoľníkov do skupiny fajčiarov a nefajčiarov za predpokladu rôznych typov elipsoidálneho šumu.

## Reference

- [1] Bi J. and Zhang T. (2004) Support Vector Classification with Input Data Uncertainty. *Advances in Neural Information Processing systems (NIPS '04)* **17**, 161–168.
- [2] Sturm J.F. (1995) Using SEDUMI 1.02, a Matlab\*toolbox for Optimization over symmetric cones. (Updated for Version 1.05). *Optimization Methods and Software* **11**, 625–653.

- [3] Moguerza J.M. and Muñoz A. (2006) Support Vector Machines with Applications. *Statistical Science* **21**, 322–336.
- [4] Bhattacharyya Ch. (2004) Robust Classification of noisy data using Second Order Cone Programming approach. *Proceedings International Conference on Intelligent Sensing and Information Processing* 433–438.
- [5] Shivaswamy P.K., Bhattacharyya Ch. and Smola A.J. (2006) Second Order Cone Programming Approach for Handling Missing and Uncertain Data. *Journal of Machine Learning Research* **7**, 1283–1314.

## Dienstbier Jan

### Tail modelling in linear models by quantile regression

TUL, KAP, Voroněžská 1329/13, CZ–406 01 Liberec

jan.dienstbier@tul.cz

We deal with the description of a variable of primary interest in terms of covariates from the extreme value point of view (see [2] for an introduction to extreme value theory) – we are interested in estimating of the condition indices, extreme conditional quantiles and small exceedance probabilities. These tasks are motivated by real case studies e.g. in geology or finance. In the literature it has been proposed to use regression quantiles as a suitable covariate dependent high threshold or even base extreme value analysis on the regression quantile process, see [1]. We broaden older results such as [5] by presenting strong approximations of high regression quantiles. Hence following the steps of [4] we obtain a whole class of consistent and asymptotically normal estimators of extreme value index as Hadamard differentiable location and scale invariant functionals of the regression quantile process. As a consequence we also gain estimators of extreme conditional quantiles and small exceedance probabilities. In the contribution we concentrate on computational aspects of the problem (one can find the background theoretical results in the forthcoming dissertation thesis of the author [3]). We illustrate the properties of the estimators on simulations and Aon Re Belgium fire portfolio data, see [1]. We also discuss various ways how to construct estimators of extreme value index based on regression quantiles in respect of the theoretical results and compare them.

## Reference

- [1] Beirlant J.et al. (2004) *Statistics of Extremes, Theory and application*. John Wiley & Sons, Chichester.

- [2] de Haan L., Ferreira A. (2006) *Extreme Value Theory, An Introduction*. Springer, New York.
- [3] Dienstbier J. (2010) *Stochastical inference in the model of extreme events* doctoral thesis, MFF UK, (to appear).
- [4] Drees H. (1998) On Smooth Statistical Tail Functionals, *Scandinavian Journal of Statistics* **25**, 187–210,
- [5] Jurečková J. (1999) Regression rank scores tests against heavy-tailed alternatives, *Bernoulli* **5**, 659–676,

## Dohnal Gejza

### Modely sousledných událostí

FS ČVUT, CQR, Karlovo nám, 13, CZ–120 00 Praha 2

dohnal@nipax.cz

V příspěvku bude prezentována analýza sousledných událostí. Tyto události jsou někdy označovány jako katastrofické. Jedná se o silně (kauzálně) závislé události, které nastávají náhodně v čase, jsou neopakovatelné a jejich důsledky jsou nevratné v relativně krátkém časovém horizontu. Události se řetězovitě šíří, tento řetězec se může náhodně větvit.

V příspěvku bude popsáno několik modelů šíření sousledných událostí, které tvoří důležitou součást analýzy rizika. Ačkoli řetězec událostí vykazuje silnou závislost, lze se na něj z určitého hlediska dívat jako na markovský řetězec a spočítat některé důležité charakteristiky, jako je například rozdělení doby prvního průchodu určitým uzlem, rozdělení doby do úplné „destrukce“, pravděpodobnost „zasažení“ vybraného uzlu a podobně. Na základě těchto charakteristik lze poté například vytvářet plány na obnovu, takzvané *Disaster Recovery Plans*.

## Došlá Šárka

### Estimation of parameters of a clipped MA(1) process

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

dosla@karlin.mff.cuni.cz

Let  $\{X_t, t \in \mathbb{Z}\}$  be a sequence of independent identically distributed random variables with an absolutely continuous CDF  $F$ . Let  $a > 0$  and  $c \in \mathbb{R}$  be some constants. We consider a sequence of 0-1 valued variables  $\{\xi_t, t \in \mathbb{Z}\}$  obtained by clipping an MA(1) process  $X_t - aX_{t-1}$  at the level  $c$ , i.e.  $\xi_t = I[X_t - aX_{t-1} < c]$  for all  $t \in \mathbb{Z}$ . Such discretization by a threshold of a continuous-valued process is a common phenomenon in biology, engineering and other

areas. We deal with the estimation problem in this model. Properties of the estimators of the parameters  $a$  and  $c$ , the success probability  $p$  and 1-lag autocorrelation  $r_1$  are investigated. A numerical study is provided as an illustration of the theoretical results.

## Dvořák Marek

### On testing changes in parameters of an autoregressive model

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

`dvorak@karlin.mff.cuni.cz`

My contribution will deal with the problem of testing a single change in variance of the  $p$ -th order autoregressive process,  $AR(p)$ , at an unknown change point  $\tau$ . We propose a test based on maximum likelihood principle for detecting such type of change, find asymptotic distribution of the test statistic and compare it with the other known tests for detecting changes in variance via empirical power functions of the test statistics.

## Fabián Zdeněk

### Míry a váhy

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

`zdenek@cs.cas.cz`

Svět pravděpodobnostních měr je pestrý a rozmanitý. Zato metrik ve výběrovém prostoru je jako šafránu. Ve své přednášce připomenu metodu zavedení skalární skórové inferenční funkce pro dané rozdělení, kterou jsem v minulých přednáškách škol ROBUST použil pro konstrukci skórových momentů a jejich odhadů jakožto charakteristik výběrů z daného rozdělení. Pomocí skalární skórové funkce lze zavést relevantní metriku ve výběrovém prostoru a její derivace představuje váhu, kterou daná míra připisuje prostoru, na kterém je definována.

## Reference

- [1] Fabián Z. (2008) New measures of central tendency and variability of continuous distributions, *Comm. Statist. Theory Methods* **37**, 159–174.
- [2] Fabián Z. (2009) Confidence intervals for a new characteristic of central tendency of distributions. *Comm. Statist. Theory Methods* **38**, 1804–1814.

**Filová Lenka****Minimal efficiency of designs in quadratic regression on  $q$ -dimensional cube**

filova@fmph.uniba.sk

FMFI UK, KAMS, Mlynská dolina, SK--84248 Bratislava

Consider a saturated quadratic regression model on  $q$ -dimensional cube  $\langle -1, 1 \rangle^q$ . Using the results of Heiligers [3], we can restrict our search for optimal designs to the designs invariant with respect to the cube's symmetry group, which is generated by permutations of coordinates and reflections along the axes. We find the  $E_k$ -optimal designs ([2]), and hence we investigate the minimal efficiency of admissible designs in this model. Next, we find the maximin efficient design, and efficiencies of well-known criteria of D, E, and A-optimality. Our results suggest that A-optimal design is fairly robust under all orthogonally invariant criteria.

**Reference**

- [1] Galil Z., Kiefer J. (1977) Comparison of design for quadratic regression on cubes. *JSPI* **1**, 121–132.
- [2] Harman R. (2004) Minimal efficiency of designs under the class of orthogonally invariant information criteria. *Metrika* **60**, 137–153.
- [3] Heiligers B. (1992) Admissible experimental designs in multiple polynomial regression, *JSPI* **31**, 219–233.
- [4] Kiefer J. (1960) Optimum design in regression problems, *AMS*, 298–325.
- [5] Pukelsheim F. (1993) Optimal design of experiments. Wiley, New York.
- [6] Pukelsheim F., Studden W.J. (1993) E-optimal designs for polynomial regression. *AS* **21**, 402–415.

**Franc Jiří****Robustifikované instrumentální proměnné**

KM FJFI ČVUT, Břehová 7, CZ–11000 Praha 1

jiri.franc@centrum.cz

Klasické odhady koeficientů v regresním modelu, jako metoda nejmenších čtverců, jsou velmi citlivé na odlehlá pozorování a nemusí být konzistentní pokud není splněna podmínka ortogonality. Robustní verze instrumentálních proměnných je uvedena a rozšířena. Jsou navrženy a odůvodněny tři různé iterační algoritmy pro výpočet odhadu pomocí instrumentálních vážených

proměnných. Dále je provedeno několik simulací, pro různé druhy modelů s různým nastavením počtu odlehklých pozorování či korelace.

## **Frcalová Blažena**

### **Časoprostorové bodové procesy**

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

frcalova@karlin.mff.cuni.cz

Časoprostorové bodové procesy mají uplatnění v různých aplikacích, např. neurofyziologii, kde pomáhají modelovat elektrické impulsy v mozku (spiky). Na reálných datech získaných z neuronu krysy pohybující se v aréně a hledající potravu jsou odhadnuty charakteristiky experimentu. Budou prezentovány dva přístupy. První přístup je znám z literatury a je založen na rekurzivních rovnicích. Druhý přístup používá bodový shot-noise Cox proces.

## **Friesl Michal**

### **Konzistence neparametrického bayesovského odhadu**

FAV ZČU, KM, Univerzitní 22, CZ–306 14 Plzeň

friesl@kma.zcu.cz

Konzistence bayesovských odhadů nemusí být v případě neparametrických bayesovských odhadů, kdy parametr je nekonečněrozměrný, automaticky zaručena. Připomeneme si, jak je tomu s konzistencí aposteriorního rozdělení a neparametrických bayesovských odhadů funkce spolehlivosti, a podíváme se na konzistenci odhadu v modelu s proporcionálním cenzorováním, prezentovaného na předchozích Robustech.

## **Grendár Marian**

### **Empirická vierohodnosť**

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

marian.grendar@savba.sk

Pod značkou „empirická vierohodnosť“ sa skrývajú dve súvisiace metódy. Jednou je metóda na konštrukciu konfidenčných intervalov a testov, ktorá kombinuje všeobecnosť neparametrického prístupu s efektívnosťou vierohodnosti. Druhou je metóda na získavanie bodových odhadov pomocou maximalizácie empirickej vierohodnosti, v rámci modelu špecifikovaného prostredníctvom odhadovacích rovníc. Obe metódy sa dajú chápať ako špeciálny prípad metódy zovšeobecného minimálneho kontrastu. Zameriame sa na

otázku volby kontrastu a na existenční problémy celého přístupu, založeného na empirických odhadovacích rovnicích.

## Helisová Kateřina

### Silová mozaika jako nástroj pro odhad parametrů v modelu náhodné množiny

KM FE ČVUT, Technická 2, CZ-166 27 Praha 6

helisova@math.feld.cvut.cz

Uvažujme náhodnou množinu  $\mathbf{X}$  danou sjednocením kruhů se středy náhodně rozmístěnými v  $S \subset \mathbf{R}^2$ , libovolnými (náhodnými nebo deterministickými) poloměry a možnými vzájemnými interakcemi. Předpokládejme, že pravděpodobnostní míra této náhodné množiny je daná hustotou vzhledem k pravděpodobnostní míře nějakého Booleovského modelu, tj. náhodného procesu kruhů bez jakýchkoliv interakcí, a že je tato hustota ve tvaru

$$f(\mathbf{x}) = \frac{1}{c_\theta} (\theta_1 A_{\mathbf{x}} + \theta_2 L_{\mathbf{x}} + \theta_3 \chi_{\mathbf{x}})$$

pro libovolnou konfiguraci  $\mathbf{x} = (x_1, \dots, x_n)$  kruhů  $x_1, \dots, x_n$ , přičemž  $\theta = (\theta_1, \theta_2, \theta_3)$  je vektor parametrů,  $c_\theta$  je normalizační konstanta,  $A_{\mathbf{x}}$  značí plochu,  $L_{\mathbf{x}}$  obvod a  $\chi_{\mathbf{x}}$  Euler-Poincarého charakteristiku pro sjednocení kruhů konfigurace  $\mathbf{x}$  (tzv. Quermassův interakční model).

Otázkou je, jak při zadaných datech odhadnout parametry tohoto modelu. Jednou z možností je metoda maximální věrohodnosti s využitím MCMC simulací (viz [4] a [5]). Další způsob (zkoumán v [1]) pak využívá tzv. integrální charakterizaci Gibbsova procesu (viz [2] a [6]).

Jelikož obě tyto metody jsou v praxi výpočetně náročné, je zde využita tzv. silová mozaika popsána detailně v [3], která rozdělí sjednocení kruhů na disjunktní konvexní podmnožiny a umožňuje tak provádět pouze lokální výpočty, což činí výpočty odhadů znatelně rychlejšími.

Popis tohoto nástroje a jeho využití bude předmětem tohoto příspěvku.

## Reference

- [1] Dereudre D., Helisová K., Lavancier F. (2010) Estimating parameters in Quermass-interaction process. Přípravuje se.
- [2] Georgii H.-O. (1976) Canonical and grand canonical Gibbs states for continuum systems. *Communications of Mathematical Physics* **48**, 31–51.
- [3] Møller J., Helisová K. (2008) Power diagrams and interaction processes for unions of discs. *Advances in Applied Probability* **40**, 321–347.



- [4] Møller J., Helisová K. (2009) Likelihood inference for interacting discs. *Scandinavian Journal of Statistics*, to appear.
- [5] Møller J., Waagepetersen R.P. (2004) *Statistical Inference and Simulation for Spatial Point Processes*. Chapman and Hall/CRC, Boca Raton.
- [6] Nguyen X.X., Zessin H. (1979) Integral and differential characterizations of Gibbs processes. *Mathematische Nachrichten* **88**, 105–115.

## Hlávka Zdeněk

### O neparametrickém odhadu polohy maxima

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

hlavka@karlin.mff.cuni.cz

Odhad maxima funkce a jeho polohy bývá často zajímavější a důležitější, než odhad celé neznámé regresní funkce. V článku [2] jsou podobné odhady a jejich asymptotické vlastnosti odvozeny od standardních neparametrických jádrových odhadů [1].

Cílem příspěvku je zejména poskytnout základní přehled o neparametrických odhadech polohy maxima a některých problémech, se kterými se můžeme setkat při jejich použití. Budeme se zabývat zejména vlivem volby hodnot nezávisle proměnné na asymptotický rozptyl odhadu polohy maxima a ukážeme si, že vhodný návrh experimentu může být zvolen velice jednoduše pomocí řešení Eulerovy diferenciální rovnice [3]. Vlastnosti výsledného odhadu budou vyšetřeny v krátké simulační studii.

## Reference

- [1] Gasser T. & Müller H.-G. (1984) Estimating regression functions and their derivatives by the kernel method, *Scand. J. Statist.* **11**, 171–185.
- [2] Müller, H.-G. (1985) Kernel estimators of zeros and of location and size of extrema or regression functions, *Scand. J. Statist.* **12**, 221–232.
- [3] Smirnow W.I. (1958) *Lehrgang der höheren Mathematik, Teil IV*. VEB Deutscher Verlag der Wissenschaften, Berlin.

**Hlubinka Daniel a Kvitkovičová Andrea****Odhady a testy pro parametry lineárně unášeného Wienerova procesu**

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

hlubinka@karlin.mff.cuni.cz

V našem příspěvku se budeme věnovat lineárně unášenému Wienerovu procesu, tedy spojitému procesu  $X = (at + bW_t, t \geq 0)$ , kde  $W$  je standardní Wienerův proces. Naším hlavním cílem je porovnat odhady a testy o parametrech  $a$  a  $b$  založené v jednom případě na pozorování procesu  $X$  v deterministických časových okamžicích  $t_1 < t_2 < \dots < t_n$  a ve druhém případě na pozorováních markovských časů prvního vstupu procesu  $X$  do úrovní  $A_1 < A_2 < \dots < A_n$ .

Oba přístupy jsou založeny na nezávislých nestejně rozdělených náhodných veličinách. V obou případech se ukáže, že pro parametr  $a$  stačí jediné pozorování (to poslední) a na ostatních nezáleží. Odhady a testy jsou tím lepší, čím je větší  $t_n$ , případně  $A_n$ . Pro parametr  $b$  je naopak úplně jedno, jak velké jsou hodnoty  $t_n$  a  $A_n$ , ale důležité je, aby  $n$  bylo co největší.

V přístupu založeném na přímém pozorování je jasné, že doba trvání experimentu je právě  $t_n$ . Pro nepřímé pozorování je délka pozorování náhodná veličina. Jak si ukážeme, nemusí touto náhodnou veličinou být čas vstupu do úrovně  $A_n$ , ale experiment lze v odůvodněných případech ukončit i dříve.

**Hornišová Klára****Neparametrická kalibrácia**Ústav merania, Slovenská akadémia vied, Dúbravská cesta 9,  
SK-841 04 Bratislava

umerhorn@savba.sk

Porovnanie viacerých neparametrických kalibračných oblastí.

**Reference**

- [1] Faugeras O. P. (2009) A quantile-copula approach to conditional density estimation, *J. Multivar. An.* **100**, 2083–2099.
- [2] Gruet M.-A. (1996) A nonparametric calibration analysis, *Ann. St.* **24**, 1474–1492.
- [3] Huet S. et al. (2004) *Statistical tools for nonlinear regression. A practical guide with S-PLUS and R examples, 2nd ed.* Springer, New York.
- [4] Misquitta, P. P. (2000) Some results in non-parametric calibration. M.S. Thesis, Texas Tech Univ., Lubbock.

- [5] Misquitta P., Ruymgaart F.H. (2005) Some results on nonparametric calibration, *Comm. in St. - Theory and Methods* **34**, 1605–1616.

## Hron Karel

### Robustní metody pro kompoziční data

PřF UPOL, KMAAM. třída 17. listopadu 12, CZ–771 46 Olomouc

hron@inf.upol.cz

Kompoziční data jsou kladné vektory, jejichž složky reprezentují relativní příspěvky různých částí na celku [1]. Formálněji,  $D$ -složková kompozice  $\mathbf{x} = (x_1, \dots, x_D)$  je definována jako vektor, kde jediná relevantní informace je obsažena v podílech mezi jeho složkami. Toto vyjádření znamená, že pro  $a$  kladné nesou  $(x_1, \dots, x_D)$  a  $(ax_1, \dots, ax_D)$  v podstatě tutéž informaci a jsou tedy nerozlišitelné. Každá kompozice takto indukuje třídu ekvivalentních kompozičních vektorů. Pro zjednodušení použití kompozic tyto často reprezentujeme v tzv. uzavřené formě, tj. jako kladné vektory, součet jejich složek je roven nějaké kladné konstantě  $\kappa$ , zvolené libovolně. Obvykle pokládáme  $\kappa$  rovno 1 nebo 100 a potom tedy složky kompozic představují procentuální podíly; proto se kompoziční data často interpretují jako data procentuální nebo *data s konstantním součtem složek*, i když je výše uvedená definice zřejmě obecnější. Množina všech  $D$ -složkových kompozic  $\mathbf{x} = (x_1, \dots, x_D)$  při dané konstantě  $\kappa$  tvoří výběrový prostor na simplexu (dimenze  $D - 1$ ),

$$\mathcal{S}^D = \{\mathbf{x} = (x_1, \dots, x_D), \quad x_i > 0, \quad i = 1, \dots, D, \quad \sum_{i=1}^D x_i = \kappa\},$$

s vlastní algebraicko-geometrickou strukturou (Hilbertovým prostorem), založenou na operacích perturbace, mocninná transformace a na tzv. Aitchisonově skalárním součinu [3]. Vzhledem ke speciální geometrii na simplexu postrádá přímá aplikace standardních statistických metod na kompoziční data smysl a vede často k nerozumným výsledkům. Východiskem se zdá být použití vhodné tzv. logratio transformace (additive, centred, isometric), které zobrazují kompozice ze simplexu do euklidovského reálného prostoru [1, 2]. Zde již mohou být standardní mnohorozměrné techniky jako metoda hlavních komponent, korelační, faktorová nebo diskriminační analýza po určitých úpravách aplikovány. V případě výskytu odlehlých hodnot v datovém souboru lze potom v případě regulárních logratio transformací (additive, isometric) uvedené metody též robustifikovat [4, 5, 6, 7]. Vzhledem

k relativní škále kompozic je přitom správná identifikace odlehlých hodnot klíčem k relevantním výsledkům dané statistické analýzy. Práci s kompozičními daty ve statistickém softwaru R se věnují knihovny `compositions` a `robCompositions`; druhá z nich obsahuje i robustní verze výše uvedených statistických metod.

## Reference

- [1] J. Aitchison (1986) *The Statistical Analysis of Compositional Data*. Chapman & Hall, London.
- [2] Egozcue V. et al. (2003) Isometric logratio transformations for compositional data analysis. *Mathematical Geology* **35**, 279–300.
- [3] J. J. Egozcue, V. Pawlowsky-Glahn (2006). Simplicial geometry for compositional data. In: A. Buccianti et al., editors, *Compositional Data Analysis in the Geosciences: From Theory to Practice*, pp. 67–77, Geological Society: London.
- [4] Filzmoser P., Hron K. (2008) Outlier detection for compositional data using robust methods. *Mathematical Geosciences* **40**, 233–248.
- [5] Filzmoser P., Hron K. (2009) Correlation analysis for compositional data. *Mathematical Geosciences*, v tisku.
- [6] Filzmoser P., Hron K., Reimann C. (2009a) PCA for compositional data with outliers. *Environmetrics*, v tisku.
- [7] Filzmoser P. et al. (2009b) Robust factor analysis for compositional data. *Computers & Geosciences* **35**, 1854–1861.

## Hušková Marie

### Change point detection in trending regression

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

huskova@karlin.mff.cuni.cz

The talk will concern test procedures for detection of changes in regression models with “trending regressors”. Particularly, we consider the linear model with a change after an unknown time point  $m$ :

$$Y_{in} = h^T(i/n)\beta + h^T(i/n)\delta_n I\{i > m_n\} + e_i, \quad i = 1, \dots, n, \quad (1)$$

where  $m_n (\leq n)$ ,  $\beta = (\beta_1, \dots, \beta_p)^T$  and  $\delta_n = (\delta_{1n}, \dots, \delta_{pn})^T \neq 0$  are unknown parameters,  $h^T(t) = (h_1(t), \dots, h_p(t))^T$ ,  $h_1(t) = 1$ ,  $i = 1, \dots, n$ ,  $t \in [0, 1]$  and  $h_j(\cdot)$ ,  $j = 2, \dots, p$ , are smooth functions on  $[0, 1]$ . Finally,  $e_1, \dots, e_n$  are random errors fulfilling regularity conditions.

Theoretical part will focus on behavior of test statistics whose limit distribution belongs to extreme value type. Approximations to critical values based on limit null distribution and bootstrap will be discussed. Results of a simulation study together with application to the real data set (air carrier traffic data) will be presented.

The talk is based on the paper: *Segmenting mean-nonstationary time series via trending regressions* written jointly with A. Aue and L. Horváth.

## Magdalena Hykšová

### Filozofické interpretace pravděpodobnosti v díle

T.G. Masaryka a K. Vorovky

ÚAM FD ČVUT, Na Florenci 25, CZ–110 00 Praha 1

hyksova@fd.cvut.cz

Príspevek je venovaný filozofickému pojetí pravděpodobnosti v díle dvou originálních českých myslitelů, kteří jsou ve všeobecném povědomí zapsáni ve zcela odlišných souvislostech: Tomáše Garriguea Masaryka (1850–1937) a Karla Vorovky (1879–1929).

T. G. Masaryka, prvního československého prezidenta, jistě není třeba představovat. Všeobecně však není příliš známo, že když byl v roce 1882 jmenován mimořádným profesorem filozofie na české univerzitě v Praze, zvolil si pro svou inaugurační přednášku téma *Humova skepse a počet pravděpodobnosti*, které pak dále rozvinul ve stejnojmenném spise [1] z roku 1883; o rok později vyšla stručnější a poněkud upravená německá verze [2]. I když se ostatní Masarykovy práce týkaly především filozofie, sociologie a později také politiky, projevil zde mimořádnou znalost vývoje teorie pravděpodobnosti, zejména v souvislosti s induktivní logikou. Zmíněné spisy jsou odpovědí na Humeovu myšlenku, že závěry neúplné indukce jsou výlučně založené na zvyku, a protože idea kauzální souvislosti neodpovídá žádnému dojmu ani vnější ani vnitřní zkušenosti, je to pojem zcela bezobsažný. Masaryk zdůrazňuje logický význam počtu pravděpodobnosti a jeho potenciál pro vyvrácení tzv. Humeovy skepse, kterou charakterizuje takto: *Pouze matematika zasluhuje naši důvěru, empirické vědy jsou nejisté, protože nám uniká poznání kauzálních souvislostí fakt; neboť o empirických faktech bychom mohli získat bezpečné poznatky pouze na základě evidentního vztahu mezi příčinou a účinkem.* ([1], str. 24)

K. Vorovka, filozof a matematik, Masarykově optimismu oponoval a na rozdíl od něj se domníval, že Humeovy námitky jsou oprávněné a teorií pravděpodobnosti je nelze vyvrátit. V letech 1913–1914 Vorovka publikoval

dvojici pojednání [3] a [4], v nichž kritizoval filozofické interpretace pravděpodobnosti a poukázal na základní problém logické interpretace, kterým je určení apriorních pravděpodobností v Bayesově vzorci pro pravděpodobnost určité hypotézy, podmíněnou daným pozorováním či zkušeností.

## Reference

- [1] Masaryk T. G. (1883) *Humova skepse a počet pravděpodobnosti*. J. Otto, Praha.
- [2] Masaryk T. G. (1884) *Dav. Hume's Skepsis und die Wahrscheinlichkeitsrechnung*. Carl Konegen, Wien.
- [3] Vorovka K. (1913) *Filosofický dosah počtu pravděpodobnosti*. Česká mysl **14**, 17–30.
- [4] Vorovka K. (1914) *O pravděpodobnosti příčin*. ČPMF **43**, 81–93.

## Chochola Ondřej

### Sekvenční monitorování v kvantilové regresi

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

chochola@karlin.mff.cuni.cz

Príspevek pojednáva o testování regresních kvantilů při postupně přicházejících datech. Navazuje na práce Qu (2008) a Su and Xiao (2008), kde se také vyšetřuje možnost testování změny v předem specifikovaném kvantilu nebo intervalu kvantilů, kde se však jedná pouze o tzv. retrospektivní přístup. V tomto příspěvku navrhuje testovou proceduru pro tzv. online monitorování. Její chování je ilustrováno pomocí simulační studie.

## Reference

- [1] Qu Z. (2008) Testing for structural change in regression quantiles. *Journal of Econometrics* **146**, 170–184.
- [2] Su L. a Xiao Z. (2008) Testing for parametric stability in quantile regression models. *Statistics and Probability Letters* **78**, 2768–2775.

**Chvosteková Martina****Simultánne tolerančné intervaly v lineárnom regresnom modeli**

Ústav merania, Slovenská akadémia vied, Dúbravská cesta 9,  
SK – 841 04 Bratislava

chvosta@gmail.com

V príspevku sa budeme zaoberať simultánnymi tolerančnými intervalmi, ktoré sú využívané v mnohých meracích úlohách, najmä pri kalibrácii meracích zariadení v prípade opakovaného dopredu neznámeho počtu meraní na zariadení (pozri [4]). Uvedieme stručný prehľad známych metód na konštruovanie simultánnych tolerančných intervalov v lineárnej regresii s normálnymi chybami. Konkrétne spomenieme Liebermanovu-Millerovu metódu [2], Wilsonovu metódu [5], Modifikovanú Wilsonovu metódu [3], Limamovu-Thomasovu metódu [3] a Meeovu-Eberhardtovu-Reeveovu metódu [4]. Predstavíme presnú,  $(1 - \alpha)$ -oblasť spoľahlivosti pre parametre lineárneho regresného modelu s normálnymi chybami zkonštruovanú pomocou LRT pre testovanie nulovej hypotézy pre všetky parametre lineárneho regresného modelu s normálne rozdelenými chybami (pozri [1]), ktorá môže byť použitá na zkonštruovanie simultánnych tolerančných intervalov pre budúce pozorovania v lineárnom regresnom modeli.

**Reference**

- [1] Chvosteková M., Witkovský V. (2009) Exact Likelihood Ratio Test for the Parameters of the Linear Regression Model with Normal Errors. *MEASUREMENT 2009. Proceedings of the International Conference on Measurement*, Institute of Measurement Science, SAS, Bratislava, 53 – 56.
- [2] Lieberman G.J., Miller R.G. Jr. (1963) Simultaneous Tolerance Intervals in Regression. *Biometrika* **50**, 155 – 168.
- [3] Limam M.M.T., Thomas R. (1988) Simultaneous Tolerance Intervals for the Linear Regression Model. *JASA* **83**, 801 – 804.
- [4] Mee R.W., Eberhardt K.R., Reeve C.P. (1991) Calibration and Simultaneous Tolerance Intervals for Regression. *Technometrics* **33**, 211 – 219.
- [5] WILSON A.L. (1967) An Approach to Simultaneous Tolerance Intervals in Regression. *AMS* **38**, 1536 – 1540.

**Janáček Jiří****Variance odhadů plochy a délky pomocí periodických mřížek**

ÚF AV ČR, Vídeňská 1083, CZ-142 Praha 4

janacek@biomed.cas.cz

Základní geometrické charakteristiky objektů v ploše nebo v prostoru mohou být odhadnuty počítáním průsečíků s náhodně umístěnou periodickou mřížkou složenou z bodů, čar nebo ploch [1].

Ergodický průměr variačního koeficientu odhadu objemu je úměrný ploše povrchu tělesa, což bylo dokázáno již v [2] pro ryze konvexní tělesa v  $d$ -rozměrném prostoru s  $6d$ -hladkou opěrnou funkcí, nebo nověji pro tělesa s  $C^{1.5}$  hladkou hranicí [3].

Podobné výsledky nebyly dosud publikovány pro odhady povrchu a délky. Variance odhadů délky čar nebo povrchu ploch pomocí náhodně orientované mřížky z ploch nebo čar se skládá ze složky určené jejich vzájemnou orientací a z reziduální složky. První složka může být vypočtena z rozdělení úhlů mezi normálami zmíněných ploch a mezi tečnami přímk. Reziduální složku je snadné spočítat pro odhad povrchu konvexních těles, např. koule nebo disku, mřížkami složenými z přímk. Lze odvodit i asymptotické vzorce pro reziduální složku odhadu délky pomocí prostorových mřížek složených z rovin nebo sfér.

**Reference**

- [1] Barbier J.E. (1860) Note sur problème de l'aiguille et le jeu du joint couvert. *Journal de Mathématiques Pures et Appliquées* **5**, 273–287.
- [2] Hlawka E. (1950) Über Integrale auf konvexen Körpern I. *Monatsh für Math* **54**, 1–36.
- [3] Janáček J. (2008) An asymptotics of variance of the lattice points count. *Czechoslovak Mathematical Journal* **58**, 751–75.



## Janková Mária

### Intervaly spoľahlivosti pre spoločnú strednú hodnotu porovnanie dvoch metód

Ústav merania, Slovenská akadémia vied, Dúbravská cesta 9,  
SK-841 04 Bratislava

majka.jankova@gmail.com

V metroológii sa často stretávame s problémom stanovenia spoločnej strednej hodnoty. V praxi ide o stanovenie čo najpresnejšieho odhadu skutočnej hodnoty meranej veličiny, pričom tento odhad sa nazýva kľúčová porovnávací referenčná hodnota (KCRV - key comparison reference value). Pre jej určenie sú k dispozícii dáta z viacerých laboratórií. Chyba pozorovaní poskytnutých každým laboratóriom pritom pozostáva z tzv. laboratórnej chyby, ktorá je pre všetky pozorovania z jedného laboratória rovnaká, a z chyby jednotlivých meraní. My budeme uvažovať situáciu s rovnomerným a normálnym rozdelením laboratórnej chyby.

V tomto príspevku sa budeme zaoberať intervalovými odhadmi spoločnej strednej hodnoty. Porovnáme dve metódy intervalového odhadu: metódu založenú na metrologickom prístupe navrhnutú Witkovským a Wimmerom v [2] a zovšeobecnené intervaly (GCI - generalized confidence intervals) navrhnuté Wangom a Iyerom v [4]. Využitím Monte Carlo simulácií skúmame frekven-tistické vlastnosti oboch metód.

## Reference

- [1] Fairweather W.R. (1972) A method of obtaining an exact confidence interval for the common mean of several normal populations. *Appl. Stat.* **21**, 229–233.
- [2] Witkovský V., Wimmer G. (2007) Confidence Interval for Common Mean in Interlaboratory Comparisons with Systematic Laboratory Biases. *Measurement Science Review* **7**, 64–73.
- [3] Witkovský V., Wimmer G. (2009) Estimation of the common mean and determination of the comparison reference value. *Tatra Mt. Math. Publ.* **39**, 53–60.
- [4] Wang C.M., Iyer H.K. (2006) A generalized confidence interval for a measurand in the presence of type-A and type-B uncertainties. *Measurement* **39**, 856–863.

**Jarušková Daniela****Metody pro aproximaci horních kvantilů lineární kombinace  $\chi^2$  rozdělení**

StF ČVUT, K101, Thákurova 7, CZ–166 29 Praha 6

jarus@mat.fsv.cvut.cz

V mnoha statistických úlohách se používá testová statistika ve formě lineární kombinace  $\chi^2$  rozdělení. V knize Johnson-Kotz-Balakrishnan je popsáno několik postupů, jak v obecném případě aproximovat toto rozdělení nebo alespoň aproximovat jeho chvost, což umožňuje výpočet přibližných hodnot horních kvantilů. Mezi nejspěšnější metody patří metody aproximace inverze charakteristické funkce za pomoci sedlového bodu. Nejvíce se nám osvědčila metoda navržená Fieldem, s kterou bychom účastníky Robustu rádi seznámili.

**Reference**

- [1] Fatalov V., Richter W.D. (1992) Gaussian probabilities of large deviation for fixed or increasing dimension. *Izvestiya Akademii Nauk Armenii. Matematika* **27**, 1–15.
- [2] Field Ch. (1993) Tail areas of linear combinations of chi-squares and non-central chi-squares. *J. Statist. Comput. Simul.* **45**, 243–248.
- [3] Imhof J.P. (1961) Computing the distribution of quadratic forms in normal variables. *Biometrika* **48**, 419–426.
- [4] Johnson N.L., Kotz S., Balakrishnan N. (1994) Continuous univariate distributions, Vol.1, 2nd ed. J. Wiley & Sons, New York

**Jonáš Petr****Robustní odhad vícerozměrného modelu lineární regrese**

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

petr.jonas@centrum.cz

V klasické statistice se k odhadování vícerozměrného modelu lineární regrese používá metoda nejmenších čtverců, která je velmi citlivá na odlehlá pozorování. Proto se v tomto příspěvku budeme zabývat robustními alternativami k této metodě a představíme metodu MLWS (Multivariate Least Weighted Squares), která je přímým zobecněním metody LWS (Least Weighted Squares).

**Jurczyk Tomáš****Vliv multikolinearity a odlehlých pozorování**

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

jurczyk@karlin.mff.cuni.cz

Tento příspěvek ukáže problémy regresní analýzy spojené s výskytem odlehlých pozorování a multikolinearity. Skrz grafy ztrátových funkcí je vysvětleno problémové chování klasických nejmenších čtverců, ale také robustních nejmenších useknutých čtverců (LTS). Nejzajímavější je odhalení nefunkčnosti právě zmíněných LTS v rozpoznávání odlehlých pozorování v situaci, kdy většina dat trpí multikolinearitou a zbytek dat je kontaminace. Z tohoto pak plyne i nemožnost využití LTS jako nástroje k detekci multikolinearity. V posteru je také představena metoda, která by měla být schopna vypořádat se současně jak s odlehlými pozorováními, tak s multikolinearitou. Náš návrh je logickou kombinací postupů, které se používají pro vypořádání se s odlehlými pozorováními, a postupů řešících multikolinearitu.

**Jurečková Jana****Asymptotika versus konečně mnoho pozorování**

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

jurecko@karlin.mff.cuni.cz

Rozdíly mezi chováním odhadů a jiných postupů při konečném počtu pozorování a asymptotickým chováním jsou často nejen kvantitativní, ale i kvalitativní. Proto je třeba chápat asymptotiku jako užitečnou pomůcku, ale její aplikaci je třeba uvážit případ od případu. Koneckonců, k asymptotice se uchylujeme, pokud neumíme odvodit rozdělení pravděpodobností odhadu při konečném rozsahu výběru, nebo pokud je toto příliš složité.

Např. pokud data mají rozdělení s těžkými chvosty, mají těžké chvosty i odhady parametru polohy, robustní či nikoli, i když robustní odhady mají často asymptoticky normální rozdělení, a tedy lehké chvosty. Jako takové jsou asymptoticky přípustné vzhledem ke kvadratické ztrátové funkci, ačkoli jsou často nepřípustné při konečném  $n$  pro jakoukoli hustotu rozdělení. To může nastat i pro maximálně věrohodný odhad, který je asymptoticky vydatný, ale nemusí být vydatný při konečném  $n$ . Další příklad je  $t$ -rozdělení o  $p$  stupních volnosti, které má těžké chvosty pro každé pevné  $p$ , ačkoli se rychle blíží k normálnímu rozdělení při  $p \rightarrow \infty$ . Takových příkladů je celá řada, proto přistupujme k asymptotice s opatrností. Známé jsou jednokrokové (iterační) verze implicitně definovaných odhadů parametru  $\theta$ , o kterých

je známo, že jsou asymptoticky blízké neiterovanému odhadu při každém  $\sqrt{n}$ -konsistentním počátečním odhadu  $\hat{\theta}_0$ ; ale zkusme je počítat pro různá  $\theta_0$ !

## Juríček Jozef

### Maximization of the information divergence from multinomial distributions

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

jozef.juricek@matfyz.cz

The explicit solution of the problem of maximization of information divergence from the family of multinomial distributions is presented, using result of N. Ay and A. Knauf for the problem of maximization of multi-information, which is the special case of maximization of information divergence from hierarchical models.

The problem of maximization of information divergence from an exponential family has emerged in probabilistic models for evolution and learning in neural networks that are based on infomax principles. The maximizers admit interpretation as stochastic systems with high complexity w.r.t. exponential family.

## Kalousová Anna

### Joseph Bertrand

FEL ČVUT, KM, Technická 2, CZ–166 27 Praha 6

kalous@math.feld.cvut.cz

Výuku pravděpodobnosti před druhou světovou válkou silně ovlivnila kniha Josepha Bertranda *Calcul des probabilités* (první vydání je z roku 1889) a to nejen ve Francii. Ačkoli byla vysoce ceněna významnými matematiky (Darboux, Poincaré, Borel), byla po válce kritizována kvůli příliš literárnímu stylu, ignorování nejnovějších poznatků a přílišnému omezování role matematické analýzy.

Joseph Bertrand se narodil 11. 3. 1822 v Paříži. Když mu bylo devět let, otec tragicky zemřel a Josepha se ujal jeho strýc, matematik Jean-Marie Duhamel (1797–1872). Joseph projevoval velké matematické nadání, v jedenácti začal navštěvovat přednášky na pařížské *École polytechnique*, v sedmnácti získal doktorát. Působil na *École polytechnique*, *École des mines*, *Lycée Saint-Louis*, *École normale supérieure* a na *Collège de France*. Napsal řadu učebnic, které byly mezi studenty dost oblíbené. Zabýval se také historií matematiky.

Byl přijat do francouzské *Académie des sciences* v roce 1854 a do *Académie française* v roce 1884. Od roku 1874 až do své smrti 3. 4. 1900 byl stálým tajemníkem matematické sekce *Académie des sciences*.

Joseph Bertrand se také věnoval geometrické pravděpodobnosti. V knize *Calcul des probabilités* najdeme známý Bertrandův paradox, který obrátil pozornost k základnímu požadavku na úlohy geometrické pravděpodobnosti, tj. k nezbytnosti pohybově invariantního zadání. V dvoudílném *Traité de calcul différentiel et de calcul intégral* je celá kapitola věnovaná Croftonově větě, kterou vysvětluje v širším kontextu geometrické pravděpodobnosti.

## Kaluža Jan

### Maximální nerovnost pro stochastickou konvoluci řízenou martingalem

MFF UK, KPMS, Sokolovská 83, CZ–186 75 Praha 8 a  
 ÚTIA AV ČR, Pod vodárenskou věží 4, CZ–182 08 Praha 8  
 kaluza@karlin.mff.cuni.cz

V příspěvku jsou zkoumány stochastické konvoluce s kontraktivní semigrupou, řízené lokálním  $L_2$ -martingalem v Hilbertově prostoru. Je podán důkaz maximální nerovnosti a exponenciální integrovatelnosti stochastické konvoluce, založený na unitárních dilatacích.

## Klaschka Jan

### O výpočtu Blakerova konfidenčního intervalu

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

Blakerův konfidenční interval pro parametr  $p$  binomického rozdělení [1] je jedním z řešení problému, jak konstruovat interval sice exaktní, tj. pokrývající skutečnou hodnotu parametru vždy s pravděpodobností rovnou nejméně nominální hladině spolehlivosti  $1 - \alpha$ , ale “méně konzervativní” než klasický Clopper-Pearsonův interval [3].

Meze Blakerova konfidenčního intervalu tvoří infimum a supremum množiny  $\{p; \beta(p) \geq \alpha\}$ , kde  $\beta$ , tzv. konfidenční křivka, je jistá funkce, která má na první pohled jednoduchý průběh: Je-li  $\hat{p}$  relativní četnost “úspěchů”, je  $\beta$  na  $[0, \hat{p}]$  neklesající a na  $[\hat{p}, 1]$  nerostoucí.

Potíž je v tom, že co platí na první pohled, na druhý už ne. Monotonie  $\beta$  na uvedených intervalech může být porušena, z čehož plynou při numerickém výpočtu konfidenčních mezí komplikace. V příspěvku se na ně podíváme.

Výpočetní algoritmus navržený Blakerem [1], [2] postupuje od Clopper-Pearsonovy meze a hledá, kde funkce  $\beta$  překračuje hladinu  $\alpha$ , v iteracích s konstantním krokem. To má za následek pomalou konvergenci (zvláště při velmi krátkém kroku) a riziko (rostoucí s délkou kroku), že algoritmus správnou mez mine a najde nesprávné řešení. Algoritmy s rychlejší konvergencí, jako např. půlení intervalu, však vzhledem k porušení monotonie mohou dospět k chybnému řešení ještě spíše.

V příspěvku bude popsán algoritmus, který využívá znalostí o průběhu funkce  $\beta$  a ke správnému řešení konverguje rychle a spolehlivě.

## Reference

- [1] Blaker H. (2000) *Confidence curves and improved exact confidence intervals for discrete distributions*. Canadian J. of Statistics **28**, 783–798.
- [2] Blaker H. (2001) *Corrigenda: Confidence curves and improved exact confidence intervals for discrete distributions*. Canadian J. of Statistics **29**, 681–681.
- [3] Clopper C. J., Pearson E. S. (1934) *The use of confidence or fiducial limits illustrated in the case of the binomial*. Biometrika **26**, 404–413.

## Klein Daniel, Žežula Ivan

### Orthogonal decompositions in growth curve models

P.J. Šafárik University, Jesenná 5, SK–04 Košice

daniel.klein@upjs.sk

The orthogonal decompositions in the growth curve model seem to be very useful and recently discussed in papers, especially those of Ye, Wang and Hu. They simplify many theoretical results. The orthogonal decomposition will be discussed in the basic and also extended growth curve model with regard to special variance structures.

## Komárek Arnošt

### Sdružené modelování spojitých i diskrétních longitudinálních dat s exkurzí do diskriminační a shlukové analýzy

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

arnost.komarek@mff.cuni.cz

V medicínské (ale i jiné) praxi nezdá se provádět na jednotlivých pacientech (či jiných jednotkách) opakovaná měření (často v čase) jistých ukazatelů (krevní tlak, přítomnost viru, počet replikací viru, ...), které souvisejí

s nepřímou pozorovatelnou charakteristikou pacienta, například typem jeho diagnózy, či vážností jeho stavu v blízké či vzdálenější budoucnosti. V příspěvku ukážeme, jak lze pomocí zobecněných lineárních smíšených modelů (GLMM) sdruženě modelovat vývoj několika spojitých a/nebo diskretních ukazatelů v čase a jak lze na základě odhadnutého modelu roztřídit pacienty do předem určeného počtu skupin (např. dle diagnózy). Výklad bude doplněn ukázkami na reálných datech a stručným popisem souvisejícího programového vybavení.

## Konár Ondřej

### Detekce zvýšených ztrát v distribuční síti zemního plynu

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

[konar@cs.cas.cz](mailto:konar@cs.cas.cz)

Již pátým rokem je v Ústavu informatiky AV ČR ve spolupráci s plynárenským průmyslem vyvíjen nelineární regresní model GAMMA pro odhad spotřeby zemního plynu. Součástí zpracování je i implementace a rutinní provoz v západních Čechách. V souvislosti s organizačními změnami v českém plynárenství přicházejí i nové požadavky na využití modelu. Novým úkolem je detekce zvýšených ztrát v distribuční síti. K dispozici budou data z desítek uzavřených lokalit (tj. lokalit, kde existuje pouze jeden vstup a žádný výstup kromě spotřeby koncových odběratelů). Cílem projektu je navrhnout statistický test, pomocí něhož se určí, zda jsou ztráty v dané lokalitě přiměřené či nepřiměřené (například z důvodu úniku plynu nebo černého odběru). Nedílnou součástí je zpracování nemalého množství reálných dat, zatížených velkou chybovostí. V příspěvku bude popsána současná situace (projekt je v začátcích) a zváženy možnosti dalšího vývoje.

## Kotík Lukáš

### Lokální regresní hloubka

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

[kotik@karlin.mff.cuni.cz](mailto:kotik@karlin.mff.cuni.cz)

Regresní hloubka (Rousseeuw 1999) je přirozeným zobecněním klasických jednorozměrných kvantilů pro lineární regresi a tedy umožňuje uspořádání regresních ploch podle míry jejich „centrality“. Nejhlubší (mediánová) regresní přímka slouží jako robustní odhad v klasickém regresním modelu. Dalším využitím regresní hloubky je konstrukce podmíněných kvantilů při dané hodnotě regresorů, tzv. regresních kvantilů. Tyto odhady lze např. použít při konstrukci pásu spolehlivosti pro regresní přímku.

Lokalizace regresní hloubky nás oprošťuje od předpokladu linearity a umožňuje nám získat odhad v modelu bez jakýkoliv regresních parametrů, tzv. neparametrické regresi. Podobně jako v lineárním modelu, lokální regresní hloubka umožňuje odhad regresních kvantilů, ale nyní bez jakýkoliv parametrických předpokladů (tzv. neparametrická kvantilová regrese). Poster bude zaměřen na základní vlastnosti těchto odhadů.

### **Kraus David, Panaretos Victor, Maddocks John**

#### **Inference druhého řádu pro gaussovské náhodné křivky s aplikací na geometrii minikroužků DNA**

EPFL SB IMA SMART, Station 8, CH – 1015 Lausanne

david.kraus@epfl.ch

Uvažujeme problém porovnání dvou výběrů gaussovských náhodných funkcí. Práce je motivována problémem rozhodnout, zda sekvence párů bází mají vliv na mechanické vlastnosti krátkých vláken DNA. Zaměřujeme se na testování rovnosti kovarianční struktury náhodných křivek. Navržené testy je založeny na empirickém Karhunen–Loèvoevě rozvoji a useknuté aproximaci Hilbert–Schmidtovy vzdálenosti empirických kovariančních operátorů. Metoda je použita na datový soubor tak zvaných minikroužků DNA, což jsou uzavřené křivky v trojrozměrném prostoru získané jako obrazy molekul DNA elektronovým mikroskopem. Výsledky naznačují významný vliv sekvencí párů bází na geometrické vlastnosti molekul.

### **Kulich Michal**

#### **Analýza stratifikovaných dvoufázových studií s kalibrovanými a odhadnutými vahami**

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

kulich@karlin.mff.cuni.cz

Dvoufázové studie šetří prostředky tím, že nevyžadují měření obtížně zjistitelných veličin na všech jedincích původně zahrnutých do výběru, ale pouze na jedincích náhodně vybraných do tzv. druhé fáze, kteří obvykle tvoří jen malou část původního výběru. Analýza dvoufázové studie je obvykle založena na různých modifikacích Horvitzova-Thompsonova odhadu. V tomto příspěvku ukážeme, jak zobecnit Horvitzův-Thompsonův odhad pro parametrické a semiparametrické modely s pomocí moderních výsledků z teorie výběrových šetření a asymptotické statistiky, jak tyto zobecněné odhady zlepšují asymptotické vlastnosti odhadů a jak lze tyto metody používat v praxi s pomocí knihovny *survey* systému *R*.



## Kvaszová Milena

### Jak studenti rozumějí základním statistickým pojmům

MÚ AV ČR, Žitná 25, CZ-115 67 Praha 1

milena.sp@centrum.cz

Při výuce statistiky předpokládáme, že studenti chápou základní statistické pojmy stejně jako my. Ale je tomu opravdu tak? Ve svém výzkumu jsem se zaměřila na zjištění toho, jak studenti rozumějí statistickým pojmům jako průměrný, náhoda, vzorek a proměnlivost. Jedním ze zajímavých výsledků výzkumu je jev, který můžeme označit jako „interference přirozeného jazyka“. V přirozeném jazyce často průměrný označuje nevýrazného, nezájímavého člověka a vzorek představuje malé množství výrobků, který dostaneme zdarma na vyzkoušení. Tyto odpovědi se objevují zhruba u 20% studentů. Další velmi rozšířený názor je, že náhoda neexistuje. Vždy lze najít příčinu. Náhodné události se dějí velice zřídka a nedá se do nich zasahovat a nijak je ovlivnit. Jako třetí zajímavý jev můžeme uvést snahu statistický údaj reprezentovat objektem. Průměrná velikost rodiny 2.5 představuje dva dospělé a malé dítě.

## Lachout Petr

### Aproximativní řešení a hodnota účelové funkce

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

ÚTIA AV ČR, Pod vodárenskou věží 4, CZ-182 07 Praha 8

lachout@karlin.mff.cuni.cz

S optimalizačními úlohami se setkáváme velmi často a přirozeně. Jako příklad můžeme uvést hledání optimálního řízení, nalezení eficientního statistického odhadu, maximalizace zadaného kritéria (např. zisku, užitku), minimalizace ztrát, nákladů, atd.

Často je však nalezení optimálního řešení ekonomicky či časově neúnosné. Navíc, optimální řešení sebou může přinášet nebezpečí, že malá perturbace vstupních parametrů úlohy, např. drobný zásah z venčí, může mít za následek, že řešení, které bylo před změnou optimální, se ukáže nevhodné či dokonce nepřípustné.

Po diskuzi se zadavateli se také často zjistí, že ani nemají zájem o řešení optimální, ale spíše o řešení robustní vzhledem k perturbacím a dávající vyšší zisk oproti řešení používanému v praxi.

Proto je užitečné hledat řešení, která nejsou sice optimální, ale dávají hodnotu kritéria, která dostatečně dobře aproximuje jeho hodnotu optimální. Takováto řešení v tomto příspěvku nazýváme aproximativní řešení. V zásadě

se jedná a dva typy:  $\varepsilon$ -optimální řešení a úrovnňová řešení (level-optimal solution).

Situace se komplikuje pokud řešená úloha obsahuje náhodné vlivy. Pak je několik možností, jak chápat, definovat, rozšířit definici aproximativních řešení. V tomto příspěvku představíme několik možností, jak to udělat. Představíme vztahy mezi nimi a jejich základní vlastnosti. Také se podíváme, jak pro ně fungují základní typy konvergence, jako je konvergence skoro jistě, v pravděpodobnosti a v distribuci.

## **Lechnerová Radka<sup>[1]</sup>, Lechner Tomáš<sup>[2]</sup>**

### **Aplikace bodových procesů při analýze veřejné správy v ČR**

<sup>[1]</sup> SVSEŠ, s.r.o., Lindnerova 575/1, CZ-180 00 Praha 8-Libeň,

<sup>[2]</sup> NF VŠE, katedra práva, nám. W. Churchilla 4, CZ-130 67 Praha 3

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

Prostorové rozmístění poskytovatelů veřejných služeb je důležitým indikátorem při zjišťování ukazatelů efektivity a účinnosti výkonu veřejné správy. Jestliže danou kategorii veřejné služby poskytují pouze některé z daných orgánů veřejné moci, lze prostorové rozmístění těch orgánů, které vybranou službu poskytují, účinně zkoumat za pomoci sumárních statistik bodových procesů. Použitá metoda navíc není závislá na konkrétním rozmístění orgánů v rámci území, tj. dává spolehlivé odpovědi, i když orgány nejsou rovnoměrně náhodně rozmístěné v rámci území ČR, což je realitou. V příspěvku se zabýváme prostorovým rozmístěním a vzájemnými interakcemi úřadů územních samosprávných celků v ČR, které provozují elektronické podatelny, své datové schránky si aktivovali dříve v rámci přechodného období, provozují Czech POINT, vykonávají přenesenou působnost na úseku stavebního úřadu nebo poskytují další služby v rámci výkonu přenesené působnosti. Získané výsledky podstatným způsobem doplňují obraz výkonu veřejné správy v České republice.

## **Linka Aleš, Kula Jiří, Tunák Mároš**

### **Detekce defektů plošných struktur**

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

ales.linka@tul.cz

**Maciak Matúš****Bootstrapping of M-smoothers**

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

mmatthew@matfyz.cz

This work is motivated by the article of Hwang R.C. (2004), Antoch J. et al. (2006) and Bickel P.J. and Freedman D.A. (1981) and it brings forward an idea of discontinuous and robust regression modelling approach.

Asymptotic distribution for local polynomial M-smoothers is derived and it heavily depends on some unknown quantities. However, knowledge of this distribution is crucial for a hypotheses testing in the change-point problem. Instead of using some plug-in techniques, which provide poor approximation, a bootstrap algorithm to simulate the unknown distribution is proposed and a proper justification of this algorithm in case of M-smoothers is given. A residual based bootstrap is investigated for a model with homoscedastic errors and, on the other hand, a nonparametric bootstrap procedure is studied for a model with heteroskedastic errors. The results are illustrated through a simulation study and an application to the global temperature data with some possible change-point occurrences.

**Reference**

- [1] Hwang R.-C. (2004) Local Polynomial M-smoothers in Nonparametric Regression, *Journal of Statistical Planning and Inference* **126**, 55–72.
- [2] Antoch J., Gregoire G., Hušková M. (2006) Test for Continuity of Regression Function, *Journal of Statistical Planning and Inference* **137**, 753–777.
- [3] Bickel P.J., Freedman D.A. (1981) Some asymptotic theory for the bootstrap, *The Annals of Statistics*, *9*, 1196–1217.
- [4] Maciak M. (2007) M-smoothers in Testing and Estimating, *WDS'07 Proceeding of Contributed Papers* **73/I**, 169–174.

**Madurkayová Barbora****Ratio type statistics for detection of changes in mean and the bootstrap method**

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

madurka@karlin.mff.cuni.cz

Procedures for detection of changes in mean are considered. In particular test procedures based on ratio type test statistics that are functionals of partial sums of residuals are studied.

Ratio type statistics are interesting for the fact that in order to compute such statistics there is no requirement to estimate the variance of the underlying model. Therefore they represent a suitable alternative for classical (non-ratio) statistics, most of all in cases when it is difficult to find a variance estimate with satisfactory properties.

We assume to have data obtained in ordered time points and study the null hypothesis of no change against the alternative of a change occurring at some unknown time point. We explore the possibility of applying the bootstrap method for obtaining critical values of the proposed test statistics.

**Marek Jaroslav, Kubáček Lubomír****O jednom geodetickém problému**

PřF UPOL, KMAAM. Třída 17. Listopadu 12, CZ-771 46 Olomouc

marek@inf.upol.cz

Mějme dva systémy souřadnic ve 2D. První systém je státní souřadnicový systém bodů v okolí nového zaměřovaného objektu (např. přehrada, tunel, budova). Druhý souřadnicový systém je vnitřní systém objektu. Souřadnice bodů objektu jsou určeny v tomto systému oproti státním souřadnicím mnohem přesněji. Problém je určit státní souřadnice vnitřních bodů a současně nepokazit vnitřní přesnost bodů nového objektu.

**McLoone Jon****New statistical features of Mathematica**

Wolfram Ltd.

jonm@wolfram.co.uk

The talk will discuss the design principles behind the technical computing system Mathematica. It will explain how these principles are applied in the implementation of the statistical capabilities of the system and with examples

from both the currently available version of Mathematica and the forthcoming release.

## Nagy Stanislav

### Hĺoubka funkcionálnych dát

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

s.nagy@volny.cz

Jedným zo základných neparametrických nástrojov štatistickej analýzy mnohorozmerných dát je štatistická hĺbka. Pre konečnerozmerné dáta existuje niekoľko rôznych kategórií hĺbkových funkcií ako uvádza [2], koncept hĺbky sa však na funkcionálne a špeciálne nekonečnerozmerné funkcionálne dáta doposiaľ nepodarilo úspešne zovšeobecniť.

V prípade konečnerozmerných funkcionálnych dát sa ponúka možnosť využiť izomorfizmus takéhoto priestoru s euklidovským priestorom, v prípade nekonečnerozmerných funkcionálnych dát však takýto postup zlyháva. Iným prístupom k problematike hĺbky nekonečnerozmerných dát je koncept *pásových hĺbok* popísaný v [1], ktoré je možné v konečnerozmernom prípade považovať za zovšeobecnenie simplexovej hĺbky dát.

V príspevku sa budeme zaoberať podmienkami, pri splnení ktorých je hĺbka konečnerozmerných funkcionálnych dát indukovaná konečnerozmernou hĺbkou na priestore koeficientov voči báze skutočne štatistickou hĺbkovou funkciou v zmysle [2] a na príkladoch porovnáme tieto výsledky s pásovými hĺbkami takýchto funkcií. Nakoniec načrtneme jeden možný postup zovšeobecnenia pásových hĺbok funkcií pomocou zahrnutia vlastností derivácií funkcie do výpočtu pásovej hĺbky a na príkladoch tento prístup porovnáme s predchádzajúcimi.

## Reference

- [1] Lopez-Pintado S., Romo J. (2007) Depth-based inference for functional data. *CSDA* **51**, 4957–4968.
- [2] Zuo Y., Serfling (2000) General notions of statistical depth function. *AS* **28**, 461–482.

**Navrátil Radim****Chování pořadových testů v lineárním modelu za přítomnosti chyb měření**

MFF UK, KPMS, Sokolovská 83, CZ-186 75 Praha 8  
r.navratil@centrum.cz

V lineárním modelu s chybami měření budeme zkoumat chování pořadových testů, které byly původně navrženy pro testování hypotéz v modelech, které předpokládaly, že odezva i regresory byly měřeny přesně. Ukážeme, že pro některé hypotézy se použitím těchto testů v modelech s chybami měření zachovává hladina testu, přítomnost chyb však vede ke snížení síly použitého testu. Toto budeme ilustrovat numericky i pomocí simulací.

**Novák Petr****Testy dobré shody pro model zrychleného času v analýze přežití**

MFF UK, KPMS, Sokolovská 83, CZ-186 75 Praha 8  
novakp@karlin.mff.cuni.cz

V příspěvku studujeme regresní modely pro analýzu přežití, věnujeme se především možnostem, jak sestavit testy dobré shody pro model zrychleného času. Porovnáváme je s testy pro Coxův model proporcionálního rizika založenými na teorii čítacích procesů. Na simulovaných datech zkoumáme empirické vlastnosti testů těchto modelů, pozorujeme jejich sílu a hladinu v závislosti na velikosti sledovaného výběru, typu regresorů a tvaru základního rizika. Hledáme, v jakých situacích je možné dobře rozlišit, podle kterého modelu se data chovají a naopak kdy je rozlišení mezi modely obtížnější.

**Pawlas Zbyněk****Odhad rozdělení dob mezi událostmi z krátkých časových oken**

MFF UK, KPMS, Sokolovská 83, CZ-186 75 Praha 8  
pawlas@karlin.mff.cuni.cz

Pro stacionární časový bodový proces můžeme pomocí Palmova rozdělení definovat distribuční funkci doby mezi dvěma událostmi, označme ji  $F(t)$ . Zajímá nás odhad této distribuční funkce na základě realizací nezávislých stejně rozdělených bodových procesů pozorovaných v pevném časovém okně  $[0, \Delta]$ . Přitom máme na mysli především situaci, kdy okno pozorování je krátké. Tím rozumíme, že délka okna je srovnatelná se střední dobou mezi událostmi. Každý pozorovaný proces tak obsahuje buď žádný, nebo velmi málo bodů.

Motivace pro studium tohoto problému pochází z neurofyzologie. Informace v nervovém systému je přenášena posloupností akčních potenciálů (tzv. spiků) generovaných jednotlivými neurony. Neuronovou aktivitu (časové uspořádání spiků) je možné modelovat stacionárním bodovým procesem. Předpokládejme, že cílový neuron dostává od populace okolních neuronů informaci, kterou musí v krátkém časovém okamžiku vyhodnotit a zareagovat. Jedním z důležitých faktorů pro toto vyhodnocení je rozdělení dob mezi dvěma následujícími spiky.

Protože nepozorujeme doby mezi událostmi delší než délka okna, nemůžeme bez dalších předpokladů na bodový proces odhadnout  $F(t)$  pro  $t > \Delta$ . Z tohoto důvodu se zaměříme hlavně na odhad distribuční funkce dob mezi událostmi za podmínky, že jsou menší než  $\Delta$ , tedy odhadujeme  $G(t) = \frac{F(t)}{F(\Delta)}$ . Zavedeme několik neparametrických odhadů a budeme studovat jejich vlastnosti teoreticky i na základě simulací. Rovněž provedeme srovnání s některými parametrickými odhady.

## Pešta Michal

### Konzistentné a ekvivariantné odhadovanie v modeli so závislými chybami v premenných

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

pesta@karlin.mff.cuni.cz

Nič nedokáže nahradiť možnosť, keď štatistik je s dátami priamo zviazaný ešte pred ich samotným získavaním. Tobôž, keď ich ešte vlastnoručne zbiera. Budeme analyzovať závislosť medzi dĺžkou a hmotnosťou dospelých jedincov pstruha obyčajného potočného (*Salmo trutta morpha fario*; L., 1758) žijúceho v prameňoch šumavskej Vydry. Keďže apriori nie je jednoznačne určené, či hmotnosť je závislá na dĺžke alebo opačne, klasický regresný model nemôže byť adekvátny. Naviac disturbancie prírody a chyby meraní kontaminujú obe sledované veličiny. Nakoniec nízka hladina prameňov Vydry má za následok limitácie v pohybe väčších jedincov smerom k samotným prameňom, čo vnáša istú formu závislosti medzi jednotlivými meraniami.

Uvažujeme model *chyby-v-premenných* so *závislými* chybami. V ňom odvodíme odhad ekvivariantý vzhľadom k merítku, otočeniu a zámene súradníc. Ukážeme, že tento odhad je totožný s odhadom získaným metódou *úplne najmenších štvorcov*. Odvodíme jeho konzistenciu i asymptotickú normalitu pre *slabo závislé* merania ( $\alpha$ - a  $\phi$ -mixing). Asymptotický rozptyl je však závislý na hodnotách vyšších momentov náhodných chýb, ktoré nie je možné odhadnúť, a tým nie je možné zostrojiť intervaly spoľahlivosti pre neznáme

parametre a ani testovať hypotézy nich sa týkajúce. Riešenie tohoto problému nám poskytne *blokový bootstrap s pohyblivými blokmi*, ktorého korektnosť však musíme dokázať.

## Petrásek Jakub

### Modelling with jump processes and optimal control

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

petrasek@karlin.mff.cuni.cz

This work discusses both a practical approach to modelling high frequency financial data with jump processes and the solution to the optimal consumption and portfolio problem. The modelling phase includes transforming the time to overcome nonstationarity, as well as choosing and calibrating the model. Two processes for modelling our data were proposed and their fit to the data compared. Furthermore, the optimal consumption and portfolio problem is studied. Both the optimal consumption and optimal portfolio are known, but under very restrictive assumptions. These assumptions are weakened. Finally, a numerical study of optimal consumption and portfolio choice on real data is performed, yielding noteworthy results.

## Picek Jan

### Odhady parametrů v modelu s chybami měření

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

jan.picek@tul.cz

Uvažujme lineární regresní model s chybami měření

$$y_i = \beta_0 + \beta_1^T \mathbf{x}_i + e_i, \quad i = 1, \dots, n,$$

$$\mathbf{x}_i^0 = \mathbf{x}_i + \mathbf{u}_i, \quad i = 1, \dots, n,$$

kde  $e_i$  je náhodná chyba a  $\mathbf{u}_i$  je vektor chyb měření vektoru  $\mathbf{x}_i$ , který je nepozorovatelný,  $\mathbf{x}_i^0$  označuje odpovídající pozorovaný vektor.

V příspěvku se budeme zabývat odhadem parametru  $(\beta_0, \beta_1^T)^T$  v modelu s chybami měření. Zaměříme se především na robustní metody (R- a M-odhady). Výsledky budou ilustrovány na základě simulací.



## Prášková Zuzana

### Bootstrap v RCA modelech

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

`zuzana.praskova@mff.cuni.cz`

Uvažujme autoregresní model s náhodnými koeficienty

$$X_t = \sum_{i=1}^p (\beta_i + B_{ti}) X_{t-i} + Y_t,$$

kde  $B_{ti}, i = 1, \dots, p$ , jsou náhodné složky autoregresních koeficientů a  $Y_t$  náhodné chyby. Model lze reprezentovat jako klasický autoregresní model

$$X_t = \sum_{i=1}^p \beta_i X_{t-i} + u_t$$

s chybami  $u_t$ , které vykazují podmíněnou heteroskedasticitu.

Pro odhad parametrů  $\beta_i$  lze použít standardní metodu nejmenších čtverců nebo další metody jako je metoda maximální věrohodnosti či metoda vážených nejmenších čtverců. Místo těchto odhadů lze uvažovat obecnější třídu odhadů (Schickovy odhady). Varianční matice asymptotického rozdělení takových odhadů je obvykle složitou funkcí zbývajících parametrů modelu, které je nutno také odhadovat.

Pro určení intervalů spolehlivosti Schickových odhadů parametrů  $\beta_i$  je v příspěvku jako alternativa k asymptotickému rozdělení navržena metoda bootstrap, která rovněž zohledňuje heteroskedasticitu chyb.

## Sabolová Radka

### Testy normality za přítomnosti rušivej regresie

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

`sabolova@karlin.mff.cuni.cz`

Uvažujeme lineárny model  $Y_i = \theta + \mathbf{x}_i' \boldsymbol{\beta} + \sigma e_i, i = 1, \dots, n$ , kde  $\mathbf{x}_i$  sú pevné regresory,  $\theta, \boldsymbol{\beta}$  a  $\sigma > 0$  neznáme parametre a  $e_i$  sú nezávislé rovnako rozdelené so spojitou distribučnou funkciou  $F$ . Chceme testovať hypotézu, že chyby  $e_1, \dots, e_n$  pochádzajú z normovaného normálneho rozdelenia.

Popíšeme modifikáciu známeho Shapiro-Wilkovho testu (Sen et. al, Austrian Journal of Statistics 32 (2003), No 1 & 2, 163–177), ktorá porovnáva maximálne vierohodný odhad parametra  $\sigma^2$  a jeho najlepší lineárny nevychýlený

odhad (BLUE) – tieto dva odhady sú asymptoticky ekvivalentné, ak chyby  $e_1, \dots, e_n$  sú normálne rozdelené. V príspevku sú ďalej prezentované výsledky simulácií, v ktorých sa skúmala sila tohoto testu.

## Schlesinger Pavel

### Využití Gibbsova algoritmu ve shlukovacích úlohách komputační lingvistiky

ÚFAL MFF UK, Malostranské nám. 2, CZ–11800 Praha 1

pavel.schlesinger@ufal.mff.cuni.cz

Shluková analýza je často používaná metoda při řešení úloh komputační lingvistiky, ať už se jedná např. o zkoumání podobnosti dokumentů nebo jednotlivých slov. Specifickou vlastností této shlukovací úlohy je velká mnoho-rozměrnost prostoru využívaných proměnných, která ovlivňuje výběr vhodné shlukové metody. Současným trendem v oboru je veden snahou o rozšíření množiny použitelných metod a o formulaci úlohy tak, aby se i zde mohl použít bayesovský přístup ke statistice. V příspěvku popíšeme model pro shlukování, který vychází z článku Walker a Ringger (2008) a který popisuje, jak lze podobné úlohy shlukování řešit použitím Gibbsova algoritmu.

## Shokirov Bobosharif, Klebanov Lev

### Estimating the proportion of false hypotheses in multiple testing procedure

MFF UK, KPMS, Sokolovská 83, CZ–18675 Praha 8

shokirov@karlin.mff.cuni.cz

One of the fundamental issues in multiple testing procedure is the detection of false positives - rejection of the null hypotheses when they are true. In the univariate case it is known as a type I error. There are different approaches in multivariate hypothesis testing to control this type of error by a priori given significant level. Although there are approaches with some relaxation of independence in multivariate case, most of them assume independence of the null hypotheses or the test statistic. Besides, in controlling multivariate analogue of univariate type I error, once the number of hypotheses to be tested is huge, we may observe a decrease in the test power. Under independence assumptions, this talk proposes a new approach for estimating the proportion of false discoveries-erroneously rejected null hypotheses. Giving the explicit form of the estimator, some of its properties are discussed.

## Reference

- [1] Carvajal-Rodriguez A., Una-Alvarez J., Rolan-Alvarez E., A new multitest correction (SGoF) that increases its statistical power when increasing the number of tests, *BMC Bioinformatics*. Available from <http://www.biomedcentral.com/content/pdf/1471>
- [2] Benjamini Y., Hochberg Y. (2000) On the adaptive control of the false discovery rate in multiple testing with independent statistics, *Journal of Educational and Behavioral Statistics* **25**.
- [3] Klebanov L.B. Yakovlev A. (2000) Diverse correlation structures in gene expression data and their utility in improving statistical inference, *Statistics and Probability Letters* **31**.
- [4] Wu W.B. (2008) On false discovery control under dependence, *The Annals of Statistics*, **36**.
- [5] Westfall P.H., Young S.S. (1993) Resampling-based multiple testing: Examples and Methods for p-value Adjustment, Wiley, New York.

## Šedová Michaela, Kulich Michal

### Dvoustupňové náhodné výběry ve výběrových šetřeních

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

[sedova@karlin.mff.cuni.cz](mailto:sedova@karlin.mff.cuni.cz)

V klasické teorii výběrových šetření se zpravidla zabýváme parametry, které charakterizující konečnou populaci (např. úhrn nebo průměr  $N$  pevných hodnot). Někdy je však vhodnější považovat pozorování za náhodné veličiny a zároveň brát v úvahu, že není k dispozici prostý náhodný výběr. V tomto kontextu uvažujeme dvoustupňové výběrové schéma, kde jsou nejprve vybrány domácnosti a poté je z každé domácnosti náhodně určen jeden člen a zařazen do studie. Uvádíme odhad střední hodnoty, který toto výběrové schéma zohledňuje, popisujeme jeho vlastnosti a porovnáváme ho s odhadem získaným z bernoulliovského výběru.

**Šnupárková Jana****Stochastická bilineární rovnice s frakcionálním šumem  
v nekonečné dimenzi**

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

snuparkova@karlin.mff.cuni.cz

Ukážeme použití nově vzniklého výsledku na příkladu, ve kterém se vyskytuje striktně eliptický diferenciální operátor druhého řádu. Stochastický integrál jest chápán ve Skorochodově smyslu.

**Staněk Jakub****Difúze v omezené oblasti s odrážející nebo pohlcující bariérou**

FS ČVUT, CQR, Karlovo nám, 13, CZ–120 00 Praha 2

Jakub.Stanek@fs.cvut.cz

Příspěvek se zabývá modelováním pohybu v uzavřené oblasti  $K$  pomocí difúze určené stochastickou diferenciální rovnicí  $dX_t = b(X_t)dt + \sigma(X_t)dB_t$ , přičemž oblast  $K$  je určena vztahem  $K = [x \in \mathbb{R}^n : f(x) \leq c]$ , kde  $c \in \mathbb{R}$  a  $f \in C^2(\mathbb{R}^n)$ . V příspěvku bude ukázáno, jak konstruovat difúzi v oblasti  $K$  tak, aby hranice  $S = [x \in \mathbb{R}^n : f(x) = c]$  byla odrážející, jak konstruovat difúzi s pohlcující hranicí  $S$  a jak difúzi na hranici  $S$ . V příspěvku budou rovněž uvedeny podmínky, za kterých je difúze na hranici  $S$  jednoznačně určena volbou trendového koeficientu  $b$ .

**Timková Jana****On Bernstein-von Mises theorem and survival analysis**

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

timkova@karlin.mff.cuni.cz

Although not well-known, the Bernstein-von Mises theorem (BvM) is a so-called bridge between bayesian and frequentist asymptotics. Basically, it states that under mild conditions the posterior distribution of the model parameter centered at the maximum likelihood estimator (MLE) is asymptotically equivalent to the sampling distribution of the MLE. This is a powerful tool especially when the classical asymptotics is tedious or impossible to conduct while bayesian asymptotic properties can be obtained via MCMC. However, in semiparametric setting with presence of infinite-dimensional parameters, as is e.g. Cox model for survival data, the results regarding BvM are more

difficult to establish but still not impossible. The proposed poster will give short overview of BvM results found in the survival analysis context.

### **Tuček Pavel a kol.**

#### **Design of measurement and statistical processing of magnetization analysis**

UPOL, kat. geoinformatiky, Tř. Svobody 26, CZ – 771 46 Olomouc

pavel.tucek@upol.cz

Príspevek je venovaný niekoľkým statistickým úlohám spojených s nelineárnou regresiou, ktoré sa vyskytli v materiálovom výskume na Univerzite Palackého v Olomouci. Prvou úlohou je problém získania odhadu veľkosti Debye-Wallerovho faktora, ktorá slúži ako miera usporiadania atómov v kryštale. Druhou úlohou je nájdenie neznámych parametrov hysterézy smyčky, ktorá jednoznačne charakterizuje dané nanomateriálové štruktúry a vypovedá o jejich aplikácii využiti. Tretím problémom je stanovenie optimálneho návrhu experimentu na získanie hľadaných parametrov. Optimálny plán merania umožňuje znížiť počet meraní a finančné náklady na experiment pri zachovaní presnosti odhadu parametrov získaných dosiaľ používanou metodikou.

### **Vencálek Ondřej**

#### **Klasifikace na základě hloubky bodu**

PřF UPOL, KMAAM. Třída 17. Listopadu 12, CZ – 771 46 Olomouc

ondrej.vencalek@upol.cz

Príspevek navazuje na prácu prezentovanú na Robustu 2008. Ukazuje využitie váženého poloprostorového hĺbky pri klasifikačnej úloze. To sa ukazuje vhodné najmä v prípade, keď sú jednotlivé distribúcie nekonvexné. V simulačnej štúdiu je srovnávané chovanie klasifikátora založeného na váženom poloprostorovom hĺbke, klasicky definované poloprostorové hĺbke a ďalších známych metódach.

## Víšek Jan Ámos

### $\sqrt{n}$ -Consistency of the least weighted squares under heteroscedasticity of error terms

IES FSV UK a ÚTIA AV ČR, Opletalova 26, CZ-11001 Prague 1

visek@fsv.cuni.cz

Let  $\mathcal{N}$  denote the set of all positive integers and  $R^p$  the  $p$ -dimensional Euclidean space. The linear regression model given for all  $n \in \mathcal{N}$

$$Y_i = X_i' \beta^0 + e_i = \sum_{j=1}^p X_{ij} \beta_j^0 + e_i, \quad i = 1, 2, \dots, n$$

will be considered. Further, for any  $\beta \in R^p$   $r_i(\beta) = Y_i - X_i' \beta$  denotes the  $i$ -th residual and  $r_{(h)}^2(\beta)$  stays for the  $h$ -th order statistic among the squared residuals, i.e. we have

$$r_{(1)}^2(\beta) \leq r_{(2)}^2(\beta) \leq \dots \leq r_{(n)}^2(\beta).$$

*Def. 1.* Let  $w : [0, 1] \rightarrow [0, 1]$  be a nonincreasing function with  $w(0) = 1$ . Then the solution of the extremal problem

$$\hat{\beta}^{(LWS, n, w)} = \arg \min \sum_{i=1}^n w \left( \frac{i-1}{n} \right) r_{(i)}^2(\beta)$$

will be called the *Least Weighted Squares*, see Víšek (2000).

Having discussed reasons and frequency of presence of heteroscedasticity of error terms  $e_i$ 's, we demonstrate (possibly significant) consequences of ignoring heteroscedasticity.

To be able to study properties of  $\hat{\beta}^{(LWS, n, w)}$ , it is plausible to show that  $\hat{\beta}^{(LWS, n, w)}$  is one of solutions of normal equations

$$NE_{Y, X, n}(\beta) = \sum_{i=1}^n w \left( F_{\beta}^{(n)}(|r_i(\beta)|) \right) X_i \left( Y_i - X_i' \beta \right) = 0 \quad (2)$$

where

$$F_{\beta}^{(n)}(r) = \frac{1}{n} \sum_{i=1}^n I \{ |r_i(\beta)| < r \} = \frac{1}{n} \sum_{i=1}^n I \{ |Y_i - X_i' \beta| < r \}.$$

As we allow for heteroscedasticity of error terms  $e_i$ 's (and hence also of residuals  $r_i(\beta)$ 's), it seems strange that we employ the e.d.f. of absolute values of residuals  $F_{\beta}^{(n)}(r)$  in (2).

Nevertheless, under mild conditions, the e.d.f.'s  $F_\beta^{(n)}(r)$  converge (uniformly in  $r \in R$  and  $\beta \in R^p$ ) weakly to the “mean” d.f. of absolute values of error terms, see Víšek (2010). It opens the way for proving  $\sqrt{n}$ -consistency of  $\hat{\beta}^{(LWS,n,w)}$ . Conditions and a sketch of proof will be given.

## Reference

- [1] Víšek J.Á. (2000) Regression with high breakdown point. Robust 2000, JČMF, Praha, 324–356.
- [2] Víšek J.Á. (2010) Empirical distribution function under heteroscedasticity. To appear in *Statistics*.

Research was supported by grant of GA ČR number 402/09/0557.

## Volf Petr

### On models for progression of record values

ÚTIA AV ČR, Pod Vodárenskou Věží 4, CZ–18208 Praha 8

volf@utia.cas.cz

Let us consider a sequence of random variables  $X_t$  and its sequential maxima. New record is established when  $X_{t+1} > \max\{X_1, \dots, X_t\}$ . The case of i.i.d. sequence has already been analyzed by many authors. However, in many real situations (e.g. from economics, finance, or climatology, too) the assumption of i.i.d. variables is not adequate. Therefore, a set of other models has been proposed describing the sequence of records (i. e. values, increments, times) with the aid of convenient model types, for intensity, regression, or time-series. In the contribution, after brief overview of approaches, we study possibilities to adapt such models to the development of sport records, namely in athletic events. Except records, it is reasonable to use data on the occurrence of ‘sub-record’ values and employ them for model formulation and evaluation. Our main concern is the statistical analysis of available data, model fitting problems and the selection of factors (covariates in a regression context) influencing the record process.

**Wimmer Gejza jr.****Konfidenčné oblasti pre regresné parametre v lineárnom zmiešanom modeli pre longitudinálne dáta**

Matematický ústav, Slovenská akadémia vied, Štefánikova 49, SK-814 73 Bratislava

wimmerg@mat.savba.sk

V praktických úlohách je často potrebné analyzovať údaje získané opakovanými meraniami viacerých subjektov v rôznych časových okamihoch. Jedným z primárnych cieľov analýzy takýchto dát je potom popísať spoločnú črtu všetkých pozorovaných subjektov, pričom pri ich modelovaní je vhodné uvažovať okrem týchto spoločných efektov navyše aj individuálne vplyvy jednotlivých subjektov na svoje opakované merania. Tu sa ukazuje využitie lineárneho zmiešaného modelu ako veľmi vhodný prístup k ich ďalšiemu spracovaniu. Pre získanie odhadov neznámych parametrov modelu sa vo veľkej miere využíva REML funkcia vierohodnosti. Na základe asymptotických vlastností takto získaného odhadu regresného parametra sme potom schopní skonstruovať konfidenčné oblasti pre regresný parameter modelu. Bohužiaľ, je všeobecne známou skutočnosťou, že v prípade malého počtu pozorovaní sú takto navrhnuté konfidenčné oblasti pre regresný parameter neprimerané. Mnoho autorov sa venovalo tejto problematike, pričom v záujme odstránenia nedostatkov tohto obvyklého prístupu ponúkajú jeho modifikácie. V príspevku budú podrobnejšie popísané dva modifikované prístupy konštrukcie konfidenčných oblastí pre regresný parameter lineárneho zmiešaného modelu vyplývajúce z článkov Kenward, Roger (1997) a Kenward, Roger (2009). Na základe simulačnej štúdie budú preverené ich vlastnosti a porovnané s vlastnosťami konfidenčných oblastí založených na asymptotických vlastnostiach REML odhadu regresného parametra pre longitudinálne dáta s AR(1) chybami.

**Reference**

- [1] Kenward M.G., Roger J.H. (1997) Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood, *Biometrics* **53**, 983–997.
- [2] Kenward M.G., Roger J.H. (2009) An improved approximation to the precision of fixed effects from restricted maximum likelihood. *Computational Statistics and Data Analysis*, doi:10.1016/j.csda.2008.12.013



**Witkovský Viktor****Využitie funkcie vierohodnosti na konštrukciu približných a exaktných konfidenčných oblastí**

Ústav merania, Slovenská akadémia vied, Dúbravská cesta 9,  
SK – 841 04 Bratislava

witkovsky@savba.sk

Konštrukcia konfidenčných oblastí pre parametre modelu založených na funkcii vierohodnosti, resp. funkcii podielu vierohodností, je alternatívou k viac tradičnej metóde konštrukcie konfidenčných oblastí tzv. waldovského typu, ktorá vychádza, za platnosti nutných predpokaldov, z asymptotickej normality rozdelenia odhadov metódou maximálnej vierohodnosti. Podľa [3] je práve postup založený na funkcii vierohodnosti v súlade s Fisherovým chápaním štatistického odhadu.

V príspevku budeme ilustrovať konštrukciu približných a exaktných konfidenčných oblastí pre parametre modelu založených na funkcii vierohodnosti v niektorých špeciálnych prípadoch, s poukázaním na možné využitie zovšeobecnených pivotov, resp. fiduciálneho rozdelenia parametrov: (i) pre parametre lineárneho regresného modelu, pozri napr. [2]; (ii) pre parametre jednoduchého modelu merania založeného na digitalizovaných meraniach, pozri napr. [4]; (iii) pre variančné komponenty v zmiešanom lineárnom modeli s dvomi variančnými komponentami, pozri napr. [1].

**Reference**

- [1] Arendacká B. (2008) Interval Estimators for a Variance Component in Mixed Linear Models with Two Variance Components. Dizertačná práca PhD vo vednom odbore 11-06-9 Pravdepodobnosť a matematická štatistika. Ústav merania SAV, Bratislava.
- [2] Chvosteková M., Witkovský V. (2009) Exact likelihood ratio test for the parameters of the linear regression model with normal errors. *Measurement Science Review* **9**, 1–8.
- [3] Uusipaikka E. (2008) What did Fisher mean by an estimate? Submitted to the *Electronic Journal of Statistics* arXiv:0807.3397v1 [math.ST].
- [4] Witkovský V. Wimmer G. (2009) Konfidenčné intervaly založené na digitalizovaných meraniach. ROBUST 2008, JČMF, Praha, 513–532.

*Pod'akovanie:* Práca vznikla vďaka podpore grantov VEGA 1/0077/09, VEGA 2/0019/10 a APVV LPP-0388-09.

**Zikmundová Markéta****Použití částicového filtru na odhad podmíněné intenzity Hawkesova procesu**

MFF UK, KPMS, Sokolovská 83, CZ-186 75 Praha 8  
zikmundova@karlin.mff.cuni.cz

Mějme omezenou množinu v rovině se středem v počátku. Uvnitř této množiny pozorujeme náhodný pohyb řídicí se obyčejnou stochastickou rovnicí. Na trajektorii tohoto pohybu (obecně neznámé) pozorujeme události. Příspěvek se zabývá simulací procesu událostí jako nekóvaného Hawkesova bodového procesu. Dále se použije částicový filtr (sekvenční Monte Carlo metoda) k odhadu časové a plošné podmíněné intenzity bodového procesu událostí a také k samotné rekonstrukci trajektorie stochastického pohybu.

**Žambochová Marta****Shlukování v souborech s odlehlými objekty pomocí metod k-průměrů**

FSE UJEP, KMS, Moskevská 54, CZ-400 96, Ústí nad Labem  
marta.zambochova@ujep.cz

Velká citlivost shlukování na odlehlá pozorování je skutečnost, která může záporně ovlivnit kvalitu výsledného rozdělení do shluků. Ve většině případů jsme odkázáni na vhodné předzpracování dat a případné vyloučení odlehlých objektů z dalšího zpracování. V odborné literatuře se však objevují i shlukovací metody přímo zaměřené na data obsahující odlehlé objekty. Jedním z takovýchto postupů je například dvoufázový algoritmus k-průměrů. V příspěvku je navržena varianta metody k-průměrů pracující s mrkd-stromy, která je postavena na jiném principu. Identifikace odlehlých objektů probíhá v rámci fáze předzpracování, kterou je nutno provádět i v případě, že nás odlehlé objekty nezajímají. Je to fáze organizující data do stromové struktury, která činí následující fázi shlukování velmi efektivní. Příspěvek pojednává o srovnání uvedených dvou metod.

**Žežula Ivan****Logistic, multinomial, and ordinal regression**

P.J. Šafárik University, Jesenná 5, SK-04 Košice  
ivan.zezula@upjs.sk

While linear regression model has its firm place in basic or intermediate statistics courses and textbooks, its generalizations have much worse position.

Even if nominal and ordinal data are frequently met in many applications, the sweetness of normality assumption seems to override everything. We will try to explain the basic principles of regression methods for categorical data, looking at both mathematical foundations and practical aspects.