

© 2004 by the author. All rights reserved. This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Antoch Jaromír	
<i>O simulaci řídkých jevů</i> .....	7
Arendacká Barbora	
<i>Přibližné konfidenční intervaly pre variančný komponent</i> <i>vo všeobecnom prípade modelu s dvomi komponentami</i> .....	7
Bartošová Jitka	
<i>Rozdělení příjmů domácností po roce 1990</i> .....	8
Beneš Viktor, Lechnerová Radka	
<i>Nelineární filtrování v Coxových bodových procesech</i> .....	8
Benko Michal	
<i>Common functional principal components</i> .....	9
Betinec Martin, Prchal Luboš	
<i>Poznámky k analýze ROC křivek</i> .....	9
Brabec Marek	
<i>Lognormální směsi a rozdělení velikosti atmosférických částic</i> ...	10
Cézová Eliška	
<i>Statistické řízení finančních toků</i> .....	10
Cimermanová Katarína	
<i>Aplikácia klasifikačných metód na analýzu</i> <i>vydychovaných plynov na detekciu pľúcnych chorôb</i> .....	10
Dienstbier Jan	
<i>Odhad Paretova indexu a kvantilová regrese</i> .....	11
Dohnal Gejza	
<i>Modely hromadné obsluhy pro řešení úloh spolehlivosti software</i> .	12
Dostál Petr	
<i>HARA – optimální obchodní strategie</i> .....	12
Fabián Zdeněk	
<i>Modifikovaná Raova vzdálenost</i> .....	13
Fajfrová Lucie	
<i>Testování hypotéz o parametru exponenciálního rozdělení</i> .....	13
Forbelská Marie	
<i>Parametrické odhady ROC křivek</i> .....	14
Friesl Michal	
<i>Porovnání neparametrických bayesovských odhadů</i> <i>při cenzorování</i> .....	14
Hlávka Zdeňek	
<i>Functional principal components for state price densities</i> .....	15
Hlubinka Daniel, Hurt Jan	
<i>Stochastická verze Kermackova-McKendrickova modelu epidemie</i>	16

Hornišová Klára	
<i>Aproximácia vnútornej a parametrickej krivosti nelineárnych regresných modelov bez použitia derivácií</i>	17
Horová Ivana	
<i>Neparametrické odhady ROC křivek</i>	17
Hušková Marie, Meintanis Stavros	
<i>Testy založené na empirických charakteristických funkcích</i>	17
Janáček Jiří	
<i>Detekce vláken v obrazových datech</i>	18
Jarušková Daniela	
<i>Analýza extrémů hydrologických a meteorologických řad</i>	18
Karlová Andrea	
<i>Kvantily v autoregresním modelu</i>	19
Kasal Roman	
<i>Aplikace genetický algoritmů v procesu plánování výroby ve společnosti Mitas, a. s.</i>	19
Klaschka Jan	
<i>O intervalových odhadech pravděpodobnosti, zvláště malých</i>	19
Kláštorecký Petr, Kulich Michal	
<i>A note on parameter estimation in regression models for case-cohort data</i>	20
Koláček Jan	
<i>Volba optimální šířky okna při jádrových odhadech regresní funkce</i>	21
Komárek Arnošt, Lesaffre Emmanuel	
<i>Lineární smíšený model pro dvojité intervalově cenzorovaná data se semiparametricky specifikovanými rozděleními</i>	21
Komárková Lenka, Dvořák Jiří	
<i>Průzkum vztahu firem k životnímu prostředí</i>	22
Konár Ondřej	
<i>Matematické modelování spotřeby zemního plynu domácností a maloodběratelů</i>	22
Koubková Alena	
<i>Rozdělení zpoždění odhadu bodu změny při sekvenčním přístupu</i>	23
Kraus David	
<i>Regresní modely pro značkové bodové procesy</i>	24
Kulich Michal	
<i>Odhadování percentilových křivek plicní funkce pomocí kvantilové regrese</i>	24

Kvitkovičová Andrea	
<i>Wienerov proces s posunutím; test hypotézy o parametri</i>	
<i>posunutia</i> .....	24
Lachout Petr	
<i>Odhady parametrů modelu a jejich konzistence</i> .....	25
Legát David	
<i>Statistika a spotřebitelské úvěry</i> .....	25
Lechnerová Radka, Beneš Viktor	
<i>Bayesovské Monte Carlo při filtrování bodových procesů</i> .....	26
Löster Tomáš	
<i>Využití systémů STATGRAPHICS Plus, SAS a MS Excel</i>	
<i>při analýze rozptylu</i> .....	26
Maciak Matúš, Ševčík Jaroslav	
<i>Výpočetné aspekty metody bootstrap</i> .....	27
Marek Jaroslav	
<i>Poslední kilometry jedné velmi známé výpravy</i> .....	27
Martínková Patrícia	
<i>Odhady vnitrotřídní korelace a spolehlivosti</i>	
<i>v didaktických testech</i> .....	27
Marušiaková Miriam	
<i>Application of permutation principle in multiple</i>	
<i>structural change test</i> .....	28
Maslowski Bohdan	
<i>Stochastické diferenciální rovnice jako modely reálných jevů</i> .....	29
Máša Petr	
<i>Zpracování a analýza obzvláště velkých reálných dat</i> .....	29
Nanásiová Olga	
<i>S-map and probability</i> .....	30
Neumanová Martina, Vávra Frant., Nový Pavel, Netrvalová Arnoštka	
<i>Model informačního vlivu a dezinformace</i> .....	30
Pawlas Zbyněk	
<i>Odhad rozptylu v kótovaných bodových procesech</i> .....	32
Pešta Michal	
<i>Isotonic regression in Sobolev spaces</i> .....	32
Pícek Jan	
<i>L-momenty a jejich některá zobecnění</i> .....	33
Plát Pavel	
<i>Signifikantnost regresních koeficientů určených metodou LWS</i> ..	33
Prášková Zuzana	
<i>Některé problémy detekce změn pro závislá pozorování</i> .....	34

Ranocha Pavel	
<i>Approximation of stationary density in some autoregressive models</i> .....	34
Reisnerová Soňa	
<i>Model vývoje nezaměstnanosti v čase a okresech</i> .....	35
Rencová Monika	
<i>Změny parametrů Weibullova rozdělení</i> .....	36
Saxl I., Ilucová L., Sklenička V., Svoboda M.	
<i>Problémy hodnocení struktury nanomateriálů</i> .....	36
Skalská Hana	
<i>Software pro odhady ROC křivek a AUC</i> .....	37
Staněk Jakub	
<i>Stochastická verze klasického modelu vývoje epidemie</i> .....	38
Strouhal Jan	
<i>Klasifikační stromy ve spolehlivosti software</i> .....	38
Stříž Pavel, Kloudová Jitka, Dobeš Kamil	
<i>Factors influencing enterprises' standing on the market</i> .....	39
Šimeček Petr	
<i>Gene expression data analysis in vitro toxicology</i> .....	39
Šimečková Marie	
<i>Statistical analysis of compulsive checking behavior in rodents – frailty models</i> .....	39
Špinková Milena	
<i>Didaktika pravděpodobnosti a statistiky</i> .....	40
Tvrdík Josef	
<i>Evoluční algoritmy</i> .....	41
Václavík Vladimír	
<i>Simulační studie robustnosti odhadů prahových parametrů některých rozdělení</i> .....	42
Vaněček Pavel	
<i>The Prospector – beyond classification</i> .....	42
Víšek Jan Ámos	
<i>Consistency of empirical distribution function of residuals in linear regression</i> .....	43
Volf Petr	
<i>Shlukování časových řad Poissonových náhodných veličin</i> .....	44
Wimmer Gejza	
<i>Niektoré matematicko-štatistické modely kalibrácie</i> .....	44
Witkovský Viktor	
<i>Analýza rozptylu a zmiešaný lineárny model</i> .....	45

**Abstrakty dodané po termínu**

Kovářová Milena	
<i>Změny denních maxim a minim teploty vzduchu</i> .....	46
Mihola Jiří <i>Měření multikolinearity s pomocí determinantů</i>	
<i>korelační matice</i> .....	46
Schlesinger Pavel	
<i>Křivky typu ROC jako nástroj pro měření a porovnání</i>	
<i>kvality na jevech počítačové lingvistiky</i> .....	46

**O SIMULACI ŘÍDKÝCH JEVŮ****Jaromír Antoch****MFF UK, KPMS, Sokolovská 83, 186 75 Praha 8 – Karlín***E-mail: jaromir.antoch@mff.cuni.cz*

Přednáška bude věnována použití principu výběru podle důležitosti (importance sampling) pro simulování jevů jež nastávají s pravděpodobností menší než  $10^{-6}$ . Pozornost bude soustředěna především na následující úlohy:

- Připomenutí metody „importance sampling“ a jejích vlastností.
- Volbě „optimální vychylující“ hustoty použité pro simulování.
- Stanovení počtu simulací.
- Příklady použití ukazující různá úskalí použitého přístupu.

Ukážeme, že klíčovým jednotícím nástrojem jsou zde klasická tvrzení matematické statistiky jakými je Cramérova či Gärdner-Ellisova věta o velkých odchylkách.

**PŘIBLIŽNÉ KONFIDENČNÉ INTERVALY  
PRE VARIANČNÝ KOMPONENT VO VŠEOBECNOM  
PRÍPADE MODELU S DVOMI KOMPONENTAMI****Barbora Arendacká****ÚM SAV, Dúbravská cesta 9, 841 04 Bratislava***E-mail: barendacka@gmail.com*

V príspevku sa budeme zaoberať metódami na konštrukciu približných intervalov spoľahlivosti pre variančný komponent zodpovedajúci náhodnému faktoru v zmiešanom lineárnom modeli s dvomi variančnými komponentami. Zameriame sa na všeobecnú situáciu s jediným obmedzením, že je k dispozícii viac ako iba minimálny počet pozorovaní. V takom prípade môžeme, okrem asymptotického intervalu, ktorý je ale nevhodný pre malé počty pozorovaní (pozri [2]), použiť interval odvodený Parkom a Burdickom [2], alebo interval navrhnutý Hartungom a Knappom [1]. V príspevku preskúmame vlastnosti týchto dvoch intervalov, ukážeme ako spolu súvisia Parkov-Burdickov interval a interval navrhnutý Thomasom a Hultquistom [3] (pôvodne odvodený pre nevyvážený model jednoduchého triedenia) a v akej forme je možné použiť posledne menovaný interval vo všeobecnej situácii. Správanie sa jednotlivých intervalov ilustrujeme tiež pomocou simulačnej štúdie.

## Reference

- [1] Hartung J., Knapp G. (2000). Confidence intervals for the between group variance in the unbalanced one-way random effects model of analysis of variance. *Journal of Statistical Computation and Simulation*, **65**, 311–323.
- [2] Park D.J., Burdick R.K. (2003). Performance of confidence intervals in regression models with unbalanced one-fold nested error structures. *Communications in Statistics Simulation and Computation*, **32**, 717–732.
- [3] Thomas J.D., Hultquist R.A. (1978). Interval estimation for the unbalanced case of the one-way random effects model. *The Annals of Statistics* **6** (3), 582–587.

## ROZDĚLENÍ PŘÍJMŮ DOMÁCNOSTÍ PO ROCE 1990

Jitka Bartošová

FM VŠE, Jarošovská 1117/II, 377 01 Jindřichův Hradec

*E-mail:* bartosov@fm.vse.cz

Transformace hospodářství České Republiky z plánované formy na tržní, která byla zahájena před více než deseti lety, má za následek mimo jiné také změny v rozdělení příjmů obyvatelstva. Vznik nových zdrojů příjmů sebou přináší změnu sociální struktury domácností, dochází také k posunu úrovně příjmů a k jejich diferenciaci uvnitř sociálních skupin i mezi nimi. Uvedené změny zapříčiňují narušení stability zvoleného statistického modelu. Dochází jednak ke kontaminaci modelu odlehlými hodnotami a jednak k růstu nesourodosti příjmů, které je typické pro směsi. Oby tyto jevy vedou ke zvýšení neshody empirického rozdělení s teoretickým modelem. Proto je předložený příspěvek zaměřen na odhad stupně kontaminace zvoleného statistického modelu rozdělení příjmů v jednotlivých sociálních skupinách a na rozklad směsí.

## NELINEÁRNÍ FILTROVÁNÍ V COXOVÝCH BODOVÝCH PROCESECH

<sup>1</sup>Viktor Beneš, <sup>2</sup>Radka Lechnerová

<sup>1</sup>MFF UK, KPMS, Sokolovská 83, 186 75 Praha 8 – Karlín

<sup>2</sup>Soukromá vysoká škola ekonomických studií,  
Lindnerova 575/1, 180 00 Praha 8 – Libeň

*E-mail:* <sup>1</sup>viktor.benes@mff.cuni.cz, <sup>2</sup>radka.lechnerova@svses.cz

Nelineární filtrování spočívá v rekonstrukci náhodné intenzity dvojně stochastického bodového procesu na základě jeho pozorování v konečném časovém intervalu. Úloha vede na problém určení podmíněné střední hodnoty. Vedle klasického přístupu pomocí stochastických diferenciálních rovnic je ukázáno alternativní řešení založené na bayesovské Monte Carlo metodě.



**COMMON FUNCTIONAL PRINCIPAL COMPONENTS****Michal Benko****CASE-Center for Applied Statistics and Economics,  
Humboldt-Universität zu Berlin***E-mail:* benko@wiwi.hu-berlin.de

Functional principal component analysis (FPCA) based on the Karhunen-Loève decomposition has been successfully applied in many applications, mainly for one sample problems. In this paper we consider common functional components for two sample problems. First we present a new method for estimation of functional principal components from discrete noisy data and discuss the asymptotic properties of proposed method. The obtained asymptotic results imply conditions under which the error caused by estimation from the discrete noisy data is first-order asymptotically negligible in comparison to the estimation based on a directly observed functional sample. Obtained results motivate the construction of “simple” bootstrap tests for one and two sample inference. For the two sample inference, we propose bootstrap test for testing the equality of eigenvalues, eigenfunctions, and mean functions of two functional samples. We illustrate the practical usage by a simulation study and application to the implied volatility analysis. This talk is based on joint work with Alois Kneip and Wolfgang Härdle. We gratefully acknowledge financial support by the Deutsche Forschungsgemeinschaft and the Sonderforschungsbereich 649 “Ökonomisches Risiko”.

**POZNÁMKY K ANALÝZE ROC KŘIVEK**<sup>1</sup>Martin Betinec, <sup>2</sup>Luboš Prchal<sup>1,2</sup>MFF UK, KPMS, Sokolovská 83, 186 75 Praha 8 – Karlín<sup>1</sup>FF UK, KSoc, Praha, <sup>2</sup>LSP UPS, Toulouse*E-mail:* <sup>1</sup>betinec@matfyz.cz, <sup>2</sup>prchal@karlin.mff.cuni.cz

ROC křivky patří v mnoha oblastech aplikované statistiky (medicína, biologie, strojové učení, lingvistika, bankovníctví) ke klasickým nástrojům hodnocení kvality užitých statistických metod, zejména klasifikátorů.

Príspevek bude venovaný statistické analýze ROC křivek a vybraným otázkám, jež se týkají jak jejich odhadů, tak testování hypotéz o ROC křivkách. Na reálných motivačních úlohách z oblasti lingvistiky, resp. biologie, představíme formální statistický pohled na ROC analýzu. Ukážeme přitom parametrický, neparametrický a funkcionální přístup k řešení jednotlivých úloh.

## **LOGNORMÁLNÍ SMĚSI A ROZDĚLENÍ VELIKOSTI ATMOSFÉRICKÝCH ČÁSTIC**

**Marek Brabec**

**Státní zdravotní ústav, Oddělení biostatistiky a informatiky,  
Šrobárova 48, 100 42 Praha 10**

*E-mail:* mbrabec@szu.cz

Příspěvek se zabývá modelováním „velikostního spektra“ atmosférických částic a jeho chování v čase. Často používaný lognormální model je zde výhodné rozšířit na lognormální směs s několika komponenty. Poukážeme na některá úskalí, která přitom vznikají (zejména při odhadu parametrů), ale i na možnosti interpretačně výhodného popisu velikostní distribuce a některých jejích prakticky zajímavých vlastností. Zakoncentrování empirické informace do relativně „malého“ parametrického modelu se pak ukáže jako užitečné i při sestavování modelu popisujícího vývoj v čase. Stručně se také zmíníme o některých souvislostech mezi odhadem komponent směsi a odhadem jistého typu nelineárních růstových křivek.

## **STATISTICKÉ ŘÍZENÍ FINANČNÍCH TOKŮ**

**Eliška Cézová**

**ČVUT, FS ÚTM, Karlovo nám. 13, 120 00 Praha 2**

*E-mail:* eliska\_c@email.cz

„Finanční zdraví“ podniku lze sledovat mimo jiné prostřednictvím finančních toků, neboť porušení finanční stability může podniku způsobit vážné problémy. Ukazuje se, že ke sledování jejich statistické stability lze použít nástrojů SPC, například Shewhartovy diagramy. Cílem příspěvku bude ukázat možnosti použití klasických regulačních diagramů pro včasnou detekci „problémů“ finančních toků.

## **APLIKÁCIA KLASIFIKAČNÝCH METÓD NA ANALÝZU VYDYCHOVANÝCH PLYNOV NA DETEKCIU PĽÚCNYCH CHORÔB**

**Katarína Cimermanová**

**ÚM SAV, Dúbravská cesta 9, 841 04 Bratislava**

*E-mail:* katarina.cimermanova@gmail.com

Rakovina je jednou z najčastejších príčin úmrtí v Európe a v západnom svete. Príčinou je neskorá diagnóza tohto ochorenia, nakoľko dostupné diagnostické

metódy nie sú dostatočne rýchle a presné. Horúcim tipom na detekciu niektorých typov rakoviny je analýza vydychovaných plynov. Táto nová neinvazívna diagnostická metóda sa zameriava na včasnú diagnózu a detekciu skorého stupňa rakoviny pľúc a pažeráka a vie poskytnúť rovnakú informáciu ako bežná diagnóza rakoviny analýzou krvi a moču. Vhodné analytické metódy na diagnózu sú plynná chromatografia s detekciou hmotnostným spektrom (GC-MS), hmotnostná spektrometria s protónovou prenosovou reakciou (PRT-MS), hmotnostná spektrometria s tokom selektívnych iónov v tube (SIFT-MS), laserová spektrometria a spektrometria pohyblivosti iónov [1].

Vo vydychovanom plyne sa dá detekovať 300 až 3000 rôznych chemických látok [1]. Význam pri diagnóze rakoviny majú len niektoré molekulové hmotnosti. Namerané dáta predstavujú koncentráciu chemickej látky vo vydychnutom plyne a jej vyjadrenie je v jednotke ppb (časť na milión). Údaje sa získavali od rôznych skupín obyvateľstva (chorí, chorí fajčiari, zdraví, zdravotnícky personál). Na základe dát z poskytnutej databázy je potrebné vybrať čo najlepšiu klasifikačnú (diagnostickú) metódu, ktorá bude s najväčšou presnosťou zaraďovať viacrozmerne dáta do skupín chorí a zdraví. V príspevku sa budeme zaoberať klasifikáciou dát pomocou Fisherovho lineárneho klasifikátora [2], metódou oporných bodov [3] a doprednými neurónovými sieťami [4].

## Reference

- [1] Aman A. *Breath analysis for clinical diagnosis and therapeutic monitoring*. World Scientific, Singapore, 2005.
- [2] Therien Ch.W. *Decision, estimation, and classification: and introduction to pattern recognition*. J. Wiley, New York, 1989.
- [3] Vapnik V. N. *Statistical learning theory*. J. Wiley, New York, 1998.
- [4] Hristev R.M. *The ANN Book*. GNU Public Licence, 1998.

## ODHAD PARETOVA INDEXU A KVANTILOVÁ REGRESE

Jan Dienstbier

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

*E-mail:* dienstbier@karlin.mff.cuni.cz

Odhad Paretova indexu byl v regresním případě, kdy jsou náhodné veličiny stochasticky závislé na vektoru nezávislých proměnných, studován v minulosti mnohem méně, než pro případ stejně rozdělených nezávislých veličin. V příspěvku je načrtnut jeden nový a zajímavý přístup využívající modifikace stávajících odhadů na základě kvantilové regrese. Zahrnut je také stručný úvod do teorie jak extrémních hodnot, tak i kvantilové regrese a přehled dosud

používaných metod odhadu Paretova indexu, které byly založeny na principech regresních kvantilů. Dále je krátce prezentována problematika důkazu konzistence a normality takto získaných odhadů. Použitelnost navrhnutého přístupu je demonstrována na simulacích.

## Reference

- [1] Beirlant J. et al. *Statistics of Extremes*. J. Wiley, Chichester, 2004.
- [2] Drees H. *On smooth statistical tail functionals*. Scandinavian J. of Statistics **25**, 187–210, 1998.
- [3] Koenker R. *Quantile regression*. Cambridge University Press, Cambridge, 2005.

## MODELY HROMADNÉ OBSLUHY PRO ŘEŠENÍ ÚLOH SPOLEHLIVOSTI SOFTWARE

**Gejza Dohnal**

**ČVUT, FS ÚTM, Karlovo nám. 13, 120 00 Praha 2**

*E-mail:* dohnal@nipax.cz

Většina matematických modelů spolehlivosti software předpokládá, že všechny chyby způsobující softwarové selhání jsou detekovatelné a beze zbytku opravitelné. Navíc se obvykle předpokládá, že při jejich odstraňování nejsou vnášeny další chyby (perfect debugging). Tento předpoklad lze použít v případě poměrně jednoduchých projektů (jednoduché webové aplikace, programy se stromovou architekturou). V případě složitých programových systémů se však tento požadavek ukazuje jako nereálný. Jeden z možných přístupů je aplikace Markovských modelů. Příspěvek ukazuje použití modelů hromadné obsluhy pro popis procesu odhalování a opravy chyb.

## HARA – OPTIMÁLNÍ OBCHODNÍ STRATEGIE

**Petr Dostál**

**MFF UK, KPMS, Sokolovská 83, 186 75 Praha 8 – Karlín**

*E-mail:* dostal@karlin.mff.cuni.cz

Uvažujme investora, který může investovat celkem do  $n$  akcií, kde  $n$  není příliš velké, a který se zajímá o odpověď na otázku, jaká je optimální investiční strategie obchodování s akciemi za předpokladu, že by platil transakční náklady pouze za  $i$ -tou akcií. Dále předpokládáme, že tržní ceny akcií se chovají jako  $n$ -dimenzionální geometrický Brownův pohyb a že cílem investora je maximalizovat asymptotiku očekávaného užítku počítaného z tržní ceny portfolia. Uvažujeme pouze užitkové funkce s konstantní hyperbolickou averzí vůči riziku (HARA). Pro názornost odvodíme první nenulový člen v Taylorově rozvoji funkce spojující transakční náklady a šířku „no-trade region”.

**MODIFIKOVANÁ RAOVA VZDÁLENOST****Zdeněk Fabián****ÚI AV ČR, Pod Vodárenskou Věží 2, 182 07 Praha 8***E-mail:* zdenek@uivt.cas.cz

Raovu vzdálenost mezi parametrem nulové hypotézy a parametrem odhadovaným lze v některých případech obtížně vyčíslit buď že je příliš komplikovaná nebo že příslušné skórové funkce nejsou monotónní. V příspěvku uvedeme modifikaci, která v takových případech může Raovu vzdálenost zastoupit.

**TESTOVÁNÍ HYPOTÉZ O PARAMETRU OBECNÉHO EXPONENCIÁLNÍHO ROZDĚLENÍ****Lucie Fajfrová****ÚTIA AV ČR, Pod Vodárenskou Věží 4, 182 07 Praha 8***E-mail:* luciefaj@hotmail.com

Obecným exponenciálním modelem (viz [1]) míníme soubor pravděpodobnostních měr  $\{P_{\theta,t} : \theta \in \Theta\}$  indexovaný parametrem  $\theta$  a dominovaný mírou  $\mu_t$ , jestliže odpovídající hustoty mají speciální – exponenciální – tvar:

$$p_{\theta,t}(\mathbf{x}_t) = \exp[T_t(\mathbf{x}_t)\theta - C_t(\theta)].$$

Index  $t$  označuje velikost pozorování a pochází z nějaké usměrněné množiny  $\mathcal{T}$ . Mezi exponenciální modely tudíž mohou patřit například Lévyho procesy či náhodná pole.

Příspěvek se bude zabývat testováním hypotéz o  $d$ -dimensionálním parametru  $\theta$  právě pro tyto modely. Testy jsou založené na tzv. Rényiho statistikách studovaných detailně v [2]. Tyto spolu s nejnovějšími výsledky ([3]), týkající se testování složených hypotéz, budou v příspěvku ilustrovány na příkladech.

**Reference**

- [1] Kůchler U., Sørensen M. *Exponential Families of Stochastic Processes*. Springer, Berlin, 1997.
- [2] Morales D., Pardo L., Vajda I. *Some new statistics for testing hypothesis in parametric model*. *Journal of Multivariate Analysis* **62**, 137–168, 1997.
- [3] Morales D., Pardo L., Pardo M.C., Vajda I. *Rényi statistics for testing composite hypothesis in general exponential model*. *Statistics* **38**, 133–147, 2004.

## PARAMETRICKÉ ODHADY ROC KŘIVEK

Marie Forbelská

PřF MU, KAM, Janáčkovo nám. 2a, 662 95 Brno

*E-mail:* forbel@math.muni.cz

ROC křivka je oblíbeným nástrojem pro vyhodnocování kvality binárních klasifikačních pravidel, která pomocí vícerozměrného znaku zařadí objekt do jedné ze dvou tříd. Postup klasifikace je založen na určitých předpokladech o vlastnostech klasifikovaných objektů, např. na předpokladu normality. Příspěvek bude věnován parametrickým metodám odhadů těchto křivek, především za předpokladu normálního a elipticky vrstevnicového rozdělení vektorů charakterizujících objekt.

## POROVNÁNÍ NEPARAMETRICKÝCH BAYESOVSKÝCH ODHADŮ PŘI CENZOROVÁNÍ

Michal Friesl

FAV ZČU, Univerzitní 22, 306 14 Plzeň

*E-mail:* friesl@kma.zcu.cz

Zabýváme se odhady funkce spolehlivosti ze zprava cenzorovaných dat. Používáme neparametrický bayesovský přístup, kdy je mezi apriorní modely pro neznámé rozdělení zkoumaných dat zahrnuta široká třída rozdělení najednou.

Na minulém Robustu jsme odvozovali neparametrický bayesovský odhad funkce spolehlivosti v Koziolově-Greenově modelu náhodného cenzorování. V něm se předpokládá, že intenzita poruch cenzorující veličiny je přímo úměrná intenzitě poruch veličiny zkoumané. Proti tomuto odhadu stojí odhad z obecného modelu, který s informativním cenzorováním nepočítá. V tomto příspěvku porovnáme oba odhady na datech z literatury a porovnáme také jejich bayesovská rizika.

## Reference

- [1] Ferguson T. S., Phadia E.G. *Bayesian nonparametric estimation based on censored data*. Annals Statistics **7**, 163–186, 1979.
- [2] Friesl M. *Neparametrické bayesovské odhady v Koziolově-Greenově modelu náhodného cenzorování*. Robust 2004 (J. Antoch a G. Dohnal eds.), JČMF, Praha, 2005, 93–100.
- [3] Hollander, M., Proschan, F. *Testing to determine the underlying distribution using randomly censored data*. Biometrics **35**, 393–401, 1979.
- [4] Hyde J. *Testing survival under right censoring and left truncation*. Biometrika **64**, 225–230, 1977.

- [5] Koziol J. A., Green S. B. *A Cramér-von Mises statistic for randomly censored data*. *Biometrika* **63**, 465–474, 1976.

## FUNCTIONAL PRINCIPAL COMPONENTS FOR STATE PRICE DENSITIES

Zdeněk Hlávka

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

*E-mail*: Zdenek.Hlavka@mff.cuni.cz

State price density (SPD) is an important concept in the option pricing theory bearing important information on the behaviour and expectations of the market. Prices  $C_t(K, T)$  of European options with strike price  $K$  observed at time  $t$  and expiring at time  $T$  allow to deduce the SPD in the following form:

$$f(K) = \exp\{r(T - t)\} \frac{\partial^2 C_t(K, T)}{\partial K^2},$$

where  $r$  is the risk-free interest rate.

Using nonlinear least squares, we construct an estimator of the state price density based on the observed option prices and satisfying shape constraints following from the no-arbitrage assumptions, see [1]. The method is then applied to the DAX option prices observed in years 1995–2003 and we estimate the SPD corresponding to each trading day in this period. The estimates corresponding to different maturities are then interpolated to one maturity common to all trading days and put on the same grid. Functional Principal Components [2] applied on these density estimates allow to display changes in the structure of the observed SPDs. The time series of the factor loadings allow to describe and visualize the changes in the structure of the SPDs in the observed period.

## Reference

- [1] Härdle, W. & Hlávka, Z. *Dynamics of state price densities*. Discussion Paper **2005-021**, SFB 649, Humboldt-Universität zu Berlin, 2005.
- [2] Ramsay J.O., Silverman B.W. *Functional Data Analysis*, Springer Verlag, New York, 1997.

**STOCHASTICKÁ VERZE KERMACKOVA –  
MCKENDRICKOVA MODELU EPIDEMIE**

**Daniel Hlubinka a Jan Hurt**

**MF F UK, KPMS, Sokolovská 83, 186 75 Praha 8 – Karlín**

*E-mail:* Daniel.Hlubinka@mff.cuni.cz, JanHurt@mff.cuni.cz

V našem příspěvku se zaměříme na klasický model časového průběhu epidemie známý jako Kermackův-McKendrickův. V tomto modelu je populace o pevné velikosti  $N$  rozdělena do tří částí:  $X$  jedinců potenciálně nemocných,  $Y$  jedinců právě nemocných a  $Z$  jedinců již uzdravených (nebo zemřelých, podle závažnosti onemocnění). Uvedený deterministický model se řídí soustavou diferenciálních rovnic

$$\begin{aligned}\dot{X}_t &= -\beta X_t Y_t, & X_0 &= x_0 \\ \dot{Y}_t &= \beta X_t Y_t - \gamma Y_t, & Y_0 &= y_0 \\ \dot{Z}_t &= \gamma Y_t, & Z_0 &= 0,\end{aligned}$$

kde  $\beta$  a  $\gamma$  jsou parametry dané epidemie. Tento model může být dále zobecněn tak, že parametr rychlosti přenášení infekce  $\beta$  uvažujeme jako funkci počtu jedinců již nemocí prošlých, tedy  $\beta = \beta(z)$ .

Náš příspěvek se zabývá převedením tohoto deterministického modelu na model stochastický, přičemž výše uvedený deterministický model zůstává „středním průběhem“ uvažované infekce. V tomto případě uvažujeme soustavu rovnic

$$\begin{aligned}dX(t) &= -\beta(X(t), Y(t), Z(t))X(t)Y(t)dt + X(t)\sigma(N(t))dW(t), \\ dY(t) &= \beta(X(t), Y(t), Z(t))X(t)Y(t)dt + Y(t)\sigma(N(t))dW(t) - \gamma Y(t)dt, \\ dZ(t) &= \gamma Y(t)dt + Z(t)\sigma(N(t))dW(t)\end{aligned}$$

s počátečními podmínkami  $x_0$ ,  $y_0$ ,  $0$  a pro velikost populace pak platí stochastická diferenciální rovnice

$$dN_t = N_t \sigma(N_t) dW_t, \quad N_0 = x_0 + y_0.$$

Ukážeme si existenční věty pro řešení a zejména simulace průběhu takovýchto epidemií pomocí programu Mathematica® pro různé volby parametrů a počáteční podmínky. Ukážeme si i vliv možných intervencí na průběh infekce.



## **APROXIMÁCIA VNÚTORNEJ A PARAMETRICKEJ KRIVOSTI NELINEÁRNYCH REGRESNÝCH MODELOV BEZ POUŽITIA DERIVÁCIÍ**

**Klára Hornišová**

**ÚM SAV, Dúbravská 9, 841 04 Bratislava**

*E-mail:* umerhornsavba.sk

Pre nelineárny regresný model s daným apriórnym rozdelením parametra navrhujeme miery jeho vnútornej a parametrickej krivosti, ktoré nezávisia od derivácií regresnej funkcie, ale iba od momentov apriórneho rozdelenia. Pre dostatočne diferencovateľný model sa v limitnom prípade nové miery zhodujú s tradičnými.

## **NEPARAMETRICKE ODHADY ROC KŘIVEK**

**Ivana Horová**

**PřF MU, KAM, Janáčkovo nám. 2a, 662 95 Brno**

*E-mail:* horova@math.muni.cz

ROC křivky (Receiver Operating Characteristic curves) byly poprvé použity v roce 1950 pro analýzu radiových signálů. Později našly tyto křivky uplatnění v široké škále různých vědních oborů. ROC křivka se používá k popisu výkonu diagnostického testu, který rozděluje pozorování do dvou skupin. Oblast, kde se nyní tyto křivky velmi často používají, je medicína, kde se jedná o rozhodovací proces zda „nemoc je“ nebo „nemoc není“. ROC křivky můžeme odhadnout parametrickými nebo neparametrickými metodami. Parametrickými metodami se zabývá příspěvek M. Forbelské. V tomto příspěvku se zaměříme na neparametrické metody. Mezi efektivní neparametrické metody patří jádrové odhady. Nejdříve jsou popsány jádrové odhady hustoty a distribuční funkce a na tomto základě jsou zkonstruovány ROC křivky. Získané odhady jsou porovnány s parametrickými odhady.

## **TESTY ZALOŽENÉ NA EMPIRICKÝCH CHARAKTERISTICKÝCH FUNKCÍCH**

**Marie Hušková, Stavros Meintanis**

**MFF UK, KPMS, Sokolovská 83, 186 75 Praha 8 – Karlín**

*E-mail:* huskova@karlin.mff.cuni.cz

Příspěvek se bude týkat několika testů založených na empirických charakteristických funkcích. Půjde jednak o testy spadající do oblasti detekce změny

rozdělení (tzv. change point problem), jednak o testy dobré shody za přítomnosti rušivých parametrů. Budou prezentovány teoretické výsledky spolu s výsledky simulací.

## **DETEKCE VLÁKEN V OBRAZOVÝCH DATECH**

**Jiří Janáček**

**Fyziologický ústav AV ČR, Vídeňská 1093, 142 00 Praha 4**

*E-mail:* janacek@biomed.cas.cz

Detekce vláken ve 2D a 3D obrazech kontaminovaných šumem je poměrně nesnadná úloha. Ke zvýraznění vláken lze použít filtry založené na vlastních číslech matice druhých derivací nebo na morfologické metodě „hit-or-miss“ vyladěné na hřebenové konfigurace obrazových elementů. Příspěvek je zaměřen na porovnání citlivosti a specifity obou metod a jejich praktické použití na snímky z konfokálního mikroskopu.

## **ANALÝZA EXTRÉMŮ HYDROLOGICKÝCH A METEOROLOGICKÝCH ŘAD**

**Daniela Jarušková**

**Stav. Fak. ČVUT, kat. matematiky, Thákurova 7, 166 29 Praha 6**

*E-mail:* jarus@mat.fsv.cvut.cz

Analýza chování extrémů meteorologických a hydrologických řad nabývá stále více na důležitosti. Výrazně silné deště trvající několik dní mají za následek rozsáhlé povodně. V posledních deseti letech se v České republice vyskytly dvě velmi rozsáhlé povodně - v roce 1997 na Moravě a v roce 2002 v Čechách. Obavy z podobných událostí vedou odborníky ke studiu průtokových a srážkových řad, přičemž speciální pozornost je věnována právě odhadům a predikci extrémů. Bohužel údaje o povodni v roce 2002 většinou chybí, protože měřicí zařízení byla z velké části zničena. Údaje o povodni v roce 1997 jsou částečně k dispozici a některé z nich nám byly poskytnuty Českým hydrometeorologickým ústavem k vědeckému bádání.

Náš první krok spočíval v modelování „chvostů rozdělení“ jednotlivých řad pomocí metody POT. V dalším kroku jsme hledali závislosti mezi extrémálním chováním řad. V příspěvku bychom Vás rádi seznámili s výsledky, ke kterým jsme dospěli, a s problémy, s kterými jsme se doposud při modelování extrémálního chování řad setkali.

**KVANTILY V AUTOREGRESNÍM MODELU****Andrea Karlová****MFF UK, KPMS, Sokolovská 83, 186 75 Praha 8 – Karlín***E-mail:* karlova@karlin.mff.cuni.cz

Obsahem tohoto příspěvku je popsání rozšíření teorie regresních kvantilů do lineárního modelu časových řad. Jsou zde základní definice a vlastnosti těchto kvantilů. Reálná data jsou pak modelována pomocí autoregresních kvantilů, jsou odhadnuty reálné monotonní funkce jedné proměnné, které reprezentují autoregresní koeficienty modelu časové řady.

**APLIKACE GENETICKÝ ALGORITMŮ V PROCESU PLÁNOVÁNÍ VÝROBY VE SPOLEČNOSTI MITAS****Roman Kasal****ÚIS FAME UTB, Mostní 5139, 760 01 Zlín***E-mail:* roman.kasal@centrum.cz

V rámci diplomové práce jsem se kromě běžně používaných statistických metod zabýval i genetickými algoritmy, jejichž pomocí jsem řešil rozplánování výroby ve firmě Mitas, a. s., Zlín. Pro aplikaci genetických algoritmů jsem se rozhodl z důvodu složitosti výrobního procesu výroby motoplášťů, který dokáže optimalizovat snad jen řízená metoda pokus-omyl. Genetické algoritmy právě tuto metodu více než vylepšily a jsou snad jediným uvažovaným nástrojem pro řešení velmi složitých problémů s velkým množstvím proměnných. Po vytvoření predikce výběrem nejvhodnější matematické či statistické metody a analýze výrobního procesu jsem v MS Excelu zkonstruoval model, který v potřebném směru odpovídal skutečné povaze výroby. Efektivnost této výroby jsem interpretoval jako množství vyrobených kusů s maximálním využitím kapacit, proto byla vybrána jako účelová funkce rovnice, jejíž výsledek určoval požadované množství vyrobených motoplášťů. Popíši způsob řešení a využitý nástroj Evolver od firmy Palisade. Součástí úkolu byly samozřejmě mnohé dílčí podúkoly, mezi které patřil problém určení vhodného pořadí receptur při vkládání do výroby. Úkol byl rovněž řešen pomocí evolučního algoritmu typizovaného pro „Problém obchodního cestujícího“ v Matlabu.

**O INTERVALOVÝCH ODHADECH PRAVDĚPODOBNOTÍ, ZVLÁŠTĚ MALÝCH****Jan Klaschka****ÚI AV ČR, Pod Vodárenskou Věží 2, 182 07 Praha 8***E-mail:* klaschka@cs.cas.cz

Intervalový odhad parametru  $p$  binomického rozdělení je úloha tak stará a notoricky známá, že se může zdát vyloučené, aby se o ní ještě dalo říct něco nového nebo zajímavého. Kupodivu se o ní i po r. 2000 objevují články dobrých autorů v dobrých časopisech.

V příspěvku bude nejdříve předveden jeden nápad, jak vylepšit exaktní konfidenční interval při extrémním počtu 0 nebo  $n$  úspěchů z  $n$  pokusů. Vzápětí se ukáže, že tento nápad (jakkoli některými propagovaný) není příliš dobrý. Dále budou připomenuty některé typy přibližných konfidenčních intervalů, které sice ne-představují žádné žhavé novinky, ale možná nejsou mezi statistickými praktiky tolik známé jako interval  $\hat{p} \pm \sqrt{\hat{p}(1-\hat{p})/n}$  (kde  $\hat{p}$  je relativní četnost úspěchů v  $n$  pokusech), a přitom mají mnohem lepší vlastnosti. V závěru možná přijde i dvouvýběrová úloha.

## A NOTE ON PARAMETER ESTIMATION IN REGRESSION MODELS FOR CASE-COHORT DATA

Petr Klášterecký, Michal Kulich

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

*E-mail:* klaster@karlin.mff.cuni.cz

The class of linear transformation models provides a general framework for analysing “classical” (or full-cohort) censored (time-to-event) data. With  $T$  being the failure time and  $Z$  the covariate vector, the models assume that  $H(T) = -\beta'Z + \varepsilon$ , where  $H$  is an unknown monotone transformation function,  $\varepsilon$  is a random variable with a known distribution which is independent of  $Z$  and  $\beta$  is an unknown (vector) regression parameter of interest. This class contains the well known proportional hazards and proportional odds models as special cases with  $\varepsilon$  following the extreme-value and the standard logistic distribution, respectively. A unified estimation procedure motivated by the counting process approach and martingale representation was developed by Chen et al. in [1]. A modification of the estimating equations suitable for case-cohort data was presented by Lu and Tsiatis who suggested weighting the equations with inverse selection probabilities in [2]. However, it can be shown that their method is seriously biased when applied to practically relevant datasets and new estimation procedures should therefore be developed.

## Reference

- [1] Chen K., Jin Z., Ying Z., *Semiparametric analysis of transformation models with censored data*. *Biometrika* **89**, 659–668, 2002.
- [2] Lu W., Tsiatis A.A., *Semiparametric Transformation Models for the Case-Cohort Study*. *Biometrika* (to appear).

**VOLBA OPTIMÁLNÍ ŠÍŘKY OKNA PŘI  
JÁDROVÝCH ODHADECH REGRESNÍ FUNKCE**

Jan Koláček

PřF MU, KAM, Janáčkovo nám. 2a, 662 95 Brno

*E-mail:* kolacek@math.muni.cz

V oblasti neparametrických metod odhadu regresní funkce představují metody jádrového vyhlazování jednu z nejúčinnějších vyhlazovacích technik. Kvalita jádrových odhadů regresní funkce závisí především na šířce vyhlazovacího okna, která řídí hladkost odhadu. Tento faktor nejvíce ovlivňuje výsledný odhad a jeho volba je zásadním problémem ve vyhlazovacích metodách. Cílem tohoto příspěvku je shrnout dosud známé metody pro odhad optimální šířky okna a uvést některé další možné přístupy k této problematice. Dále bude provedeno srovnání uvedených metod na simulovaných i reálných datech.

**LINEÁRNÍ SMÍŠENÝ MODEL PRO DVOJITĚ INTERVALOVĚ  
CENZOROVANÁ DATA SE SEMIPARAMETRICKY  
SPECIFIKOVANÝMI ROZDĚLENÍMI**

Arnošt Komárek, Emmanuel Lesaffre

Biostatistisch Centrum, Katholieke Universiteit Leuven, Leuven

*E-mail:* arnost.komarek@med.kuleuven.be

Standardní metody analýzy přežití předpokládají, že čas  $T$  do studované události je buď znám, anebo je pouze zprava cenzorován. K tomu, abychom zjistili, zda studovaná událost skutečně nastala, jsou však v mnoha případech (výzkum viru HIV, zubní lékařství) nutná (pravidelná) vyšetření (laboratorní analýza, kontrola zubařem). Uvažme příklad, kdy  $T$  je časem do výskytu zubního kazu, který lze spolehlivě diagnostikovat pouze (pravidelnou) kontrolou zubního lékaře. čas události  $T$  je potom vystaven tzv. *intervalovému cenzorování*. V určitých situacích je však nejen čas události, ale též okamžik – *počáteční čas*  $U$ , kdy jednotka začne přicházet do úvahy pro studovanou událost, zjišťován tímto způsobem. Je-li  $T$  časem do výskytu zubního kazu, potom čas  $U$  je zřejmě roven okamžiku prořezání daného zubu, neboť teprve od této chvíle je zub vystaven riziku zubního kazu. Čas do události (zubního kazu)  $T$  lze tedy zapsat jako  $T = U + V$ , kde  $U$  je počáteční (chronologický) čas a  $V$  (chronologický) čas události. Je-li též počáteční čas  $U$  intervalově cenzorován, mluvíme o *dvojitě intervalově cenzorovaných datech*.

V příspěvku se zaměříme na regresní analýzu dvojitě intervalově cenzorovaných dat v situaci, kdy ne všechna pozorování jsou nutně nezávislá, typicky

vyskytující se ve shlucích. Například, chceme-li u výše zmíněného zubařského příkladu studovat současně kazivost několika různých zubů, nelze pozorování na zubech patřících jednomu dítěti považovat za nezávislá. Pro analýzu navrhneme použít lineárního smíšeného modelu, přičemž pomocí náhodných efektů ošetříme shlukování dat. V kontextu cenzorovaných dat je velmi obtížné ověřit jakékoliv parametrické předpoklady týkající se rozdělení studovaných časů do události. Z tohoto důvodu navrhneme pro specifikaci rozdělení semiparametrický přístup motivovaný vyhlazováním pomocí splinů. Metoda bude ilustrována na analýze kazivosti trvalých prvních stoliček („šestek“) užitím dat sesbíraných ve vlámské části Belgie v letech 1996–2001 v rámci Signal Tandmobiel® projektu.

## PRŮZKUM VZTAHU FIREM K ŽIVOTNÍMU PROSTŘEDÍ

**Lenka Komárková, Jiří Dvořák**

**FM VŠE, Jarošovská 1117/II, 377 01 Jindřichův Hradec**

*E-mail:* komarkol@fm.vse.cz

Předmětem zkoumání je vzájemný vztah mezi přístupem vedoucích pracovníků firem k problematice životního prostředí, úsilím, jaké daná společnost této problematice věnuje a důsledky této strategie pro ekonomickou úspěšnost firmy. Všechny tyto otázky jsou sledovány z hlediska subjektivního hodnocení jednotlivých aktérů, což vede na problém statistického zpracování ordinálních dat.

## Reference

- [1] Ahmed N.U., Montagnano R.V., Naffzinger D.W. *Environmental Concerns, Effort and Impact: An Empirical Study*. Mid- American Journal of Business **18**, 61–69, 2003.
- [2] Agresti A. *Categorical Data Analysis, 2nd ed.*. J. Wiley & Sons., New Jersey, 2002.

## MATEMATICKÉ MODELOVÁNÍ SPOTŘEBY ZEMNÍHO PLYNU DOMÁCNOSTÍ A MALOODBĚRATELŮ

**Ondřej Konár**

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

*E-mail:* konar@cs.cas.cz

Pro plynárenské společnosti je velmi důležité mít přehled o aktuálním objemu prodaného plynu. Na rozdíl od velkoodběratelů a středních odběratelů, kteří jsou měření průběhově po hodinách, je spotřeba domácností a maloodběratelů zjišťována zpravidla jednou ročně odečtením plynoměru. Z důvodů

optimálního využití pracovních sil probíhají odečty plynoměřů cyklicky, tzn. že v každém měsíci je odečtena část (přibližně 1/12) zákazníků. V každém okamžiku proto existuje určitý objem plynu, který nebyl vyfakturován. Tento objem je třeba odhadnout pomocí vhodného matematického modelu.

V příspěvku bude představen nelineární regresní model *GAMMA* pro odhad objemu nevyfakturovaného plynu, vyvíjený v Ústavu informatiky AV ČR za podpory grantu GA AV ČR č. 1ET400300513. Model je využíván v rutinním provozu v Západočeské plynárenské, a.s. a jedná se o využití modelu v celé síti plynáren RWE v ČR.

## ROZDĚLENÍ ZPOŽDĚNÍ ODHADU BODU ZMĚNY PŘI SEKVENČNÍM PŘÍSTUPU

Alena Koubková

MF F UK, KPMS, Sokolovská 83, 186 75 Praha 8 – Karlín

*E-mail:* aja@matfyz.cz

Předpokládejme postupně přicházející data splňující model polohy  $Y_i = \mu_i + e_i$ , kde  $Y_i$  jsou data,  $\mu_i$  parametr polohy a  $e_i$  náhodné chyby. K dispozici máme trénovací soubor dat o velikosti  $m$ , pro který platí  $\mu_1 = \dots = \mu_m$  a v nějakém čase  $m + k^*$  může nastat změna:  $\mu_1 = \dots = \mu_m = \dots = \mu_{m+k^*-1} \neq \mu_{m+k^*} = \mu_{m+k^*+1} = \dots$ . Úkolem je odhalit změnu co nejdříve a přitom kontrolovat pravděpodobnost falešného poplachu.

Pro případ, kdy  $e_i$  jsou i.i.d. s distribuční funkcí  $F$  symetrickou kolem nuly, jejíž druhá derivace existuje v okolí nuly a  $F'(0) = f(0) > 0$ , byla dříve navržena testová statistika  $\tilde{Q}(m, k) = \sum_{i=m+1}^{m+k} \text{sign}(Y_i - \tilde{\mu}_m)$ , kde  $\tilde{\mu}_m = \text{medián}(Y_1, \dots, Y_m)$ . Chyba je detekována v čase

$$\tau(m) = \min\{k : |\tilde{Q}(m, k)|g^{-1}(m, k, \gamma) > c(\alpha)\},$$

kde  $g(m, k, \gamma)$  je hraniční funkce,  $\gamma$  doplňující parametr a  $c(\alpha)$  je kritická hodnota získaná z limitního Wienerova procesu.

Tento příspěvek je zaměřen na rozdělení času odhadu  $\tau(m)$  v závislosti na rozdělení  $F$ , velikosti a času změny. Pro  $k^* = O(m^\theta)$ , kde  $0 \leq \theta \leq \left(\frac{1-2\gamma}{2(1-\gamma)}\right)^2$  je odvozeno přímo rozdělení  $\tau(m)$  ve tvaru

$$\lim_{m \rightarrow \infty} P\left(\frac{\tau(m) - a(m)}{b(m)} \leq x\right) = \Phi(x),$$

kde  $a(m)$  a  $b(m)$  jsou blíže určené funkce parametru  $m$  a  $\Phi(x)$  je distribuční funkce standardního normálního rozdělení. Pro ostatní  $k^*$  je odvozen pouze řád zpoždění  $\tau(m) - k^*$ .

**REGRESNÍ MODELY PRO ZNAČKOVANÉ  
BODOVÉ PROCESY****David Kraus****MFF UK, KPMS, Sokolovská 83, 186 75 Praha 8 – Karlín***E-mail:* kraus@karlin.mff.cuni.cz

V příspěvku budou studovány značkové bodové procesy v čase, tedy procesy představující pozorování náhodných veličin (značek) v náhodných časových okamžicích. Pozornost je věnována regresním modelům, v nichž rozdělení značek jakož i intenzita časového procesu závisí na predikovatelných kovariátách.

**ODHADOVÁNÍ PERCENTILOVÝCH KŘIVEK PLICNÍ  
FUNKCE POMOCÍ KVANTILOVÉ REGRESE****Michal Kulich****MFF UK, KPMS, Sokolovská 83, 186 75 Praha 8 – Karlín***E-mail:* Michal.Kulich@mff.cuni.cz

Plicní funkce, zejména objem vydechnutého vzduchu za 1 sekundu ( $FEV_1$ ), je důležitým ukazatelem stavu plic u nemocných s dědičnou metabolickou poruchou zvanou cystická fibróza (CF). Plicní funkce závisí na pohlaví, věku a výšce a její standartisace není snadná. Cílem našeho projektu bylo vyvinout percentilové křivky pro  $FEV_1$  u nemocných s CF jako funkce pohlaví, věku a výšky. Použili jsme kvantilovou regresi aplikovanou na data z registru pacientů v USA. Společný vliv věku a výšky jsme modelovali pomocí součinné B-splínové báze v obou proměnných. Tento příspěvek popisuje a diskutuje detaily aplikace kvantilové regrese na podobné problémy.

**WIENEROV PROCES S POSUNUTÍM  
TESTY HYPOTÉZY O PARAMETRI POSUNUTIA****Andrea Kvitkovičová****MFF UK, KPMS, Sokolovská 83, 186 75 Praha 8 – Karlín***E-mail:* akvitkovicova@centrum.cz

Uvažujme spojitý náhodný proces  $\{M_t; t \geq 0\}$ , u kterého nesledujeme celý jeho priebeh, ale len čas, v ktorom prekročí hranicu  $A > 0$ . Na základe tejto informácie chceme testovať hypotézu

$H_0 : M_t = w_t$ , kde  $w_t$  je Wienerov proces,  
proti alternatíve

$H_1 : M_t = w_t + at$ , kde  $a > 0$  je známe.



Príspevok je zameraný na konštrukciu a vlastnosti dvoch testov založených na Neymanovej-Pearsonovej lemme.

V prvom prístupe použijeme rozdelenie náhodnej veličiny  $\tau_A = \inf\{t \geq 0; M_t \geq A\}$ , ktoré je za nulovej hypotézy známe a za alternatívnej hypotézy ho spočítame pomocou Girsanovovej vety. V druhom prístupe sa problém prevedie na test hypotézy o parametri binomického rozdelenia pre náhodnú veličinu  $B_n = \sum_{1 \leq i \leq n} I\{\max_{0 \leq t \leq T} M_t^i \geq A\}$ .

## ODHADY PARAMETRŮ MODELU A JEJICH KONZISTENCE

**Petr Lachout**

**MFF UK, KPMS, Sokolovská 83, 186 75 Praha 8 – Karlín**

*E-mail:* Petr.Lachout@mff.cuni.cz

Odhadování parametrů modelu založené na empirických pozorováních vede často na řešení vhodné optimalizační úlohy. V tomto příspěvku se budeme takovýmito odhady zabývat. Přesněji, budeme uvažovat odhady jež jsou  $\varepsilon$ -optimálním řešením nějaké optimalizační úlohy.

Cílem příspěvku je zdůraznit vlastnosti, které jsou určující pro konzistenci takto získaných odhadů. Odhlížíme od náhodného aspektu získaných pozorování a pohlížíme na ně jako na deterministické posloupnosti. Pokud jsou pozorování realizací náhodných veličin, které jsou i.i.d., silně stacionární, mixující, či svázané jiným vhodným statistickým mechanismem, pak jsou potřebné podmínky splněny s pravděpodobností jedna.

## Reference

- [1] Doukhan P. *Mixing: Properties and Examples*. Lecture Notes in Statistics **85**, Springer-Verlag, Berlin, 1994.
- [2] Hoffmann-Jørgensen J. (1994). *Probability with a View Towards to Statistics I, II*. Chapman and Hall, New York.
- [3] Jurečková J., Sen P.K. *Robust Statistical Procedures*. John Wiley & Sons, Inc., New York, 1996.
- [4] Štěpán J. (1987). *Teorie pravděpodobnosti*. Academia, Praha.
- [5] van der Vaart A.W., Wellner J.A. *Weak Convergence and Empirical Processes*. Springer, New York.

## STATISTIKA A SPOTŘEBITELSKÉ ÚVĚRY

**David Legát**

**MFF UK, KPMS, Sokolovská 83, 186 75 Praha 8 – Karlín**

*E-mail:* david.legat@gmail.cz

V mnoha oblastech prodeje se využívají statistické modely pro předpovídání chování klientů. Nejinak je tomu v případě poskytování spotřebitelských úvěrů, kde tyto modely slouží jak pro řízení rizika v podobě velkého množství nesplacených úvěrů, tak pro různé marketingové kampaně. Tento příspěvek obsahuje popis několika vybraných problémů, které se opakovaně vyskytují při vytváření požadovaných modelů, spolu s doporučeními, jak tyto problémy řešit.

### **BAYESOVSKÉ MONTE CARLO PŘI FILTROVÁNÍ BODOVÝCH PROCESŮ**

<sup>1</sup>Radka Lechnerová, <sup>2</sup>Viktor Beneš

<sup>1</sup>Soukromá vysoká škola ekonomických studií, s.r.o.,  
Lindnerova 575/1, 180 00 Praha 8 – Libeň,

<sup>2</sup>MFF UK, KPMS, Sokolovská 83, 186 75 Praha 8 – Karlín

*E-mail:* <sup>1</sup>radka.lechnerova@svses.cz, <sup>2</sup>viktor.benes@mff.cuni.cz

V úloze nelineárního filtrování Coxových bodových procesů lze užít bayesovský přístup a následně metoda MCMC. Příspěvek popisuje algoritmus řešení v situaci, kdy Coxův proces v čase je řízen semimarkovskou intenzitou generovanou Poissonovým procesem.

### **VYUŽITÍ SYSTÉMŮ STATGRAPHICS PLUS, SAS A MS EXCEL PŘI ANALÝZE ROZPTYLU**

Tomáš Löster

VŠE, kat. statistiky, nám. W. Churchilla 4, 130 67 Praha 3

*E-mail:* losterto@vse.cz

Analýza rozptylu je jedna ze základních statistických metod, která zkoumá závislost proměnných. Při výpočtech charakteristik potřebných k analýze rozptylu a následnému testu o nezávislosti proměnných je možné využít některý z dostupných softwarových produktů, například systém STATGRAPHICS Plus, SAS nebo MS Excel. Příspěvek se zabývá porovnáním uvedených systémů v dané oblasti, včetně souvisejících činností, kterými je například nalezení kritické hodnoty (kvantilu F rozdělení).

### **Reference**

- [1] Arltová M., Bílková D., Jarošová E., Pourová Z.: *Příklady k předmětu statistika A*. VŠE, Praha 2003.
- [2] Chajdiak J. *Štatistické úlohy a ich riešenie v Exceli*. Statis, Bratislava, 2005.
- [3] Jarošová E., Pecáková I. *Příklady k předmětu statistika B*. VŠE, Praha, 2000.

- [4] Marek L. a kol. *Statistika pro ekonomy aplikace*. Profesional Publishing, Praha, 2005.
- [5] Řezanková H. *Analýza kategoriálních dat*. VŠE, Praha, 2005.

## VÝPOČETNÉ ASPEKTY METÓDY BOOTSTRAP

**Matúš Maciak, Ševčík Jaroslav**

**MFF UK, KPMS, Sokolovská 83, 186 75 Praha 8 – Karlín**

*E-mail:* mmatthew@matfyz.cz

Účelom našej práce je poskytnúť stručný prehľad v oblasti využitia štatistickej resamplovej metódy BOOTSTRAP v štatistických softwaroch. K tomuto účelu sme zvolili v akademickej oblasti často používaný voľne dostupný štatistický program R a tiež jeho komerčnú verziu S-plus. Práca obsahuje niekoľko vzorových príkladov, ktoré slúžia k lepšej názornosti aplikácie bootstrapovej metódy. Okrem toho sme sa v niektorých prípadoch pokúsili simuláciami zisťovať, aký minimálny počet bootstrapových výberov je potrebný k dôveryhodnému použitiu metódy. Keďže ide o pomerne modernú, v súčasnosti veľmi obľúbenú a čoraz častejšie používanú štatistickú metódu, snažili sme sa poukázať aj na to, že by mala byť používaná s rozvahou a nie mechanicky. Je to síce metóda praktická, na druhej strane má však náročné teoretické pozadie. Práca by mala prispieť k získaniu prehľadu, ako správne využívať software pri stále častejšie používanej resamplovej metóde bootstrap.

## POSLEDNÍ KILOMETRY JEDNÉ VELMI ZNÁMÉ VÝPRAVY

**Jaroslav Marek**

**PřF UP Olomouc, Tomkova 40, 779 00 Olomouc**

*E-mail:* marek@inf.upol.cz

Cílem příspěvku je populárně demonstrovat možné použití teoretických úvah vedoucích k lepším výsledkům získaných z nepřesných měření. Práce je věnována konkrétní úloze, ve které jsou použita skutečná data dohledaná v beletrii. Pro určení odhadů parametrů jsou však použity smyšlené kovarianční matice. Práce je věnována dvouetapovému regresnímu modelu s podmínkou na parametry obou etap. Numerická studie obsahuje porovnání odhadů získaného pomocí metody nejmenších čtverců a  $H$ -optimálního odhadu.

## ODHADY VNITROTŘÍDNÍ KORELACE A SPOLEHLIVOSTI V DIDAKTICKÝCH TESTECH

**Patricia Martinková**

**ÚI AV ČR, EuroMISE, Pod Vodárenskou Věží 2, 182 07 Praha 8**

*E-mail:* pata@atrey.karlin.mff.cuni.cz

Uvažujme situaci, kdy pozorujeme nula-jedničkové proměnné  $Y_{ij}$ , kde  $Y_{ij} = 1$  v případě správného zodpovězení  $j$ -té položky  $i$ -tým studentem,  $i = 1, \dots, n$ ,  $j = 1, \dots, J$ . Dále předpokládejme, že  $i$ -tý student má úroveň znalosti rovnou  $a_i$ , kde  $a_i \sim N(0, \sigma_A^2)$ . Pravděpodobnost, že student s úrovní znalosti  $a_i$  správně zodpoví na položenou otázku nechť je dána vztahem  $p_i = P(Y_i = 1|a_i) = \frac{\exp(a_i)}{1 + \exp(a_i)}$ .

Zajímáme se o odhad koeficientu vnitrotřídní korelace  $\rho = \text{corr}(Y_{ij}, Y_{ij'})$   $\forall i, \forall j, j'$ . V článku [1] jsou zmiňovány dva odhady koeficientu  $\rho$ . Třetí odhad získáme, upravíme-li odhad Cronbachova alfa s přihlédnutím na příslušný počet položek. Čtvrtý odhad můžeme získat obdobnou úpravou odhadu navrženého v závěru článku [2]. Simulace napovídají, že čtvrtý odhad vykazuje sice nejmenší rozptyl, nicméně na rozdíl od prvních tří zmiňovaných odhadů také jisté vychýlení. V příspěvku se pokusíme o srovnání výše uvedených odhadů.

## Reference

- [1] Mak Tak K. *Analysing intraclass correlation for dichotomous variables*. Appl. Statist. **37**, 344–352, 1988.
- [2] Zvára K. *Měření reliability aneb Bacha na Cronbacha*. Informační bulletin České Statistické Společnosti **12**, 13–20, 2002.

## APPLICATION OF PERMUTATION PRINCIPLE IN MULTIPLE STRUCTURAL CHANGE TEST

Miriám Marušíaková

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

*E-mail:* maruskay@gmail.com

We consider multiple linear regression models with structural changes. We describe  $F$ -type tests for detection of changes, namely tests with the null hypothesis of no change against the alternative with  $k$  changes. Approximations to the corresponding critical values are usually obtained through the limit distribution of the test statistic under the null hypothesis. Here we propose another possibility - a method based on the application of the permutation principle. We present a number of simulation results and show that the permutation arguments provide satisfactory approximations to the critical values.

## STOCHASTICKÉ DIFERENCIÁLNÍ ROVNICE JAKO MODELY REÁLNÝCH JEVŮ

Bohdan Maslowski

MÚ AV ČR, Žitná 25, 110 00 Praha 1

*E-mail:* maslow@math.cas.cz

Hlavním cílem příspěvku je ilustrovat, především na příkladech, užitečnost stochastických diferenciálních rovnic pro modelování některých reálných jevů. Na začátku budou připomenuty potřebné základní pojmy a fakta, přičemž důraz bude kladen na to, co je motivuje. V další části pak budou předvedeny některé modely. Účelem je především zdůvodnění a motivace příslušných matematických pojmů a výsledků, tedy nastínění cesty od „praxe“ k „teorii“.

## ZPRACOVÁNÍ A ANALÝZA OBZVLÁŠTĚ VELKÝCH REÁLNÝCH DAT

Petr Máša

Adastra, Benešovská 10, 101 00 Praha 10

*E-mail:* Petr.Masa@AdastraCorp.com

Tento příspěvek se zabývá problémy při zpracování a analýze velkých dat v praxi. Celý proces analýzy dat zahrnuje široké spektrum činností. Ukazuje se, že drtivou většinu času při analýze dat zabírá především příprava dat (odhaduje se na 60-90% času). Tyto činnosti se mohou zdát jako méně významné, avšak zpravidla nejsou tak triviální a bez nich není možné analýzu provádět. Řeší se zejména problémy se získáním dat, integritou a kvalitou dat. Tím se stává každá analýza náročnější na zdroje IT i analytiky. Proto je třeba na začátku vyhodnotit, zdali přínosy dané analýzy alespoň pokryjí náklady, tedy zdali má smysl tuto analýzu provádět. Celou analýzu bychom pak měli provádět s cílem co nejlépe splnit původní cíl analýzy a zajistit co nejvyšší přínosy. Proto bychom měli nejprve pečlivě vydefinovat, co je řešený problém. To není triviální úloha. Vlastní příprava dat se pak zpravidla realizuje v databázi pomocí databázového jazyka SQL, případně jiného (pokud možno velmi silného) nástroje na přípravu dat. Samotná analýza dat také není jednoduchá. Je potřeba zvolit vhodný software, který by byl schopen naše (často velmi rozsáhlá) data rozumně zpracovat, tzn. data ve vhodné formě načíst, připravit pro statistickou proceduru a spustit analýzu na těchto datech. To nás limituje ve výběru metod, které použijeme. Také je třeba vyrovnat se s předpoklady statistických metod při jejich použití. V praxi je z časových důvodů obtížné zafixovat data a řešit problém neúměrně dlouhou

dobu. Přínosy řešení obvykle s časem velmi rychle klesají. Samozřejmě záleží na řešené úloze.

## S-MAP AND PROBABILITY

**Oľga Nánásiová**

**Slovak University of Technology, Dept. of Mathematics**

**Radlinského 11, 813 68 Bratislava**

*E-mail:* olga@math.sk

The modeling of random events, which are not measurable simultaneously, provides us with interesting results. The basic model in our approach is an orthomodular lattice  $L$  (OML). An  $s$ -map is a map from  $L^n \rightarrow [0, 1]$ , which help to us define a conditional state on  $L$ . In our presentation we show some examples of such maps and there application to the classical probability space.

## Reference

- [1] Nánásiová O. *On conditional probabilities on quantum logic*. Int. Jour. of Theor. Phys. **25**, 155–162, 1987.
- [2] Nánásiová O. *Map for simultaneous measurements for a quantum logic*. Int. Journ. of Theor. Phys. **42**, 1889–1903, 2003.
- [3] Khrennikov A., Nánásiová O. *Representation theorem of observables on a quantum system*, Preprint, 2003.

## MODEL INFORMAČNÍHO VLIVU A DEZINFORMACE

**Martina Neumanová, František Vávra, Pavel Nový,**

**Arnoštka Netrvalová**

**FAV ZČU, KIV, Univerzitní 22, 306 14 Plzeň**

*E-mail:* mneumano@kiv.czu.cz

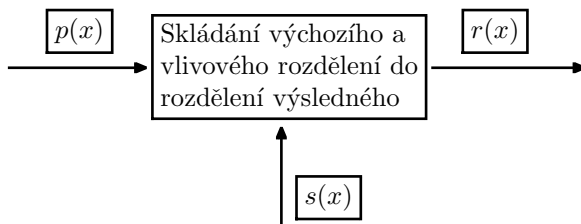
V pracích [3] a [4] jsme představili koncept dezinformace, který je založen na rozdílu mezi skutečným rozdělením pravděpodobnosti a předpokládaným modelem. Pro pojetí divergence vycházíme z „konvergence“ statistického odhadu k nějaké hodnotě. Mějme  $n$  pozorování  $\{x_1, \dots, x_n\}$ , iid. Pak pro triviální formu odhadu divergence platí:

$$\frac{1}{n} \sum_{i=1}^n \lg \frac{e(x_i)}{s(x_i)} \rightarrow \sum_{x \in X} p(x) \lg \frac{e(x)}{s(x)},$$

kde  $p(x)$  je skutečné, nedostupné, pravděpodobnostní rozdělení,  $e(x)$  je jeho model a  $s(x)$  je rozdělení, proti kterému je testováno. Označíme

$$D(p\|s; e) = \sum_{x \in X} p(x) \lg \frac{e(x)}{s(x)}$$

jako „dezinformační“ divergenci pravděpodobnostních modelů  $p(x)$  proti  $s(x)$ , přičemž je modelováno rozdělením  $e(x)$ . „Dezinformační“ divergence  $D(p\|s; e)$  je nejen užitečným statistickým pojmem, ale i dobrým prostředkem pro modelování informačního působení (reklamy, politické kampaně, ...). Situaci informačního působení můžeme, zjednodušeně, popsat následujícím schématem:



V tomto schématu rozdělení  $p(x)$  modeluje stav před informační kampaní (působením),  $s(x)$  je popisem stavu po takové kampani a  $r(x)$  je modelem vlivového působení (kampaně). Vhodnou mírou pro efekt informačního působení je klasická divergence  $D(p\|s)$ . Tu lze dekomponovat do dvou složek:

$$D(p\|s) = D(p\|r) + D(p\|s; r).$$

První člen součtu popisuje „vztah“ mezi ovlivňovaným a ovlivňujícím rozdělením. Druhý člen je modelem odlišnosti mezi ovlivňovaným a výsledným rozdělením, přičemž ovlivňované je substituováno ovlivňujícím rozdělením. Předkládaný příspěvek je věnován diskusi o „přelévání“ efektu působení mezi oběma členy součtu a jeho interpretaci.

## Reference

- [1] Vajda I. *Teória informácie a statistického rozhodovania*. Alfa Bratislava, 1982.
- [2] Cover T.M., Thomas J.A. *Elements of Information Theory*. J. Wiley, 1991.
- [3] Vávra F., Nový P., *Informace a dezinformace*. Seminář z aplikované matematiky. Katedra aplikované matematiky, Přírodovědecká fakulta MU, Brno, 13. 4. 2004.
- [4] Kotlíková M., Mašková H., Netrvalová A., Nový P., Spíralová D., Vávra F., Zmrhal D., *Informace a dezinformace – statistický pohled*. ROBUST'2004, JČMF 2004. Třešť 2004.

## ODHAD ROZPTYLU V KÓTOVANÝCH BODOVÝCH PROCESECH

Zbyněk Pawlas

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

*E-mail:* Zbynek.Pawlas@mff.cuni.cz

V prostorové statistice často pracujeme s měřeními jisté veličiny v nepravidelně rozmístěných bodech v rovině nebo prostoru. Polohy těchto bodů i naměřené hodnoty jsou považovány za náhodné. Tuto situaci můžeme modelovat pomocí kótovaných bodových procesů.

Naším cílem je studium vlastností odhadů rozptylu typické kóty. Vzhledem k tomu, že nevylučujeme korelace mezi kótami, tak klasický výběrový rozptyl je vychýlený. Pro rozsahy výběrů, které nejsou příliš velké, může být střední čtvercová chyba nezanedbatelná. Zavedeme jiný odhad rozptylu založený na  $U$ -statistice druhého řádu. Pro vybrané typy kótovaných bodových procesů porovnáme kvalitu odhadů rozptylů teoreticky i na základě simulační studie.

## ISOTONIC REGRESSION IN SOBOLEV SPACES

Michal Pešta

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

*E-mail:* michal.pesta@matfyz.cz

We propose a class of nonparametric estimators for regression models based on least squares over sufficiently smooth sets of functions. These estimators permit the imposition of additional monotonicity and concavity constraints.

Estimation takes place over balls of functions which are elements of suitable Sobolev space. The Sobolev spaces are special types of Hilbert spaces that facilitate calculation of least square projection. The Hilbertness is allowing us to take projections and hence to decompose spaces into mutually orthogonal complements. We assemble and prove necessary preliminaries and theorems for statistical regression in these spaces. Then we transform the problem of searching for the best fitting function in an infinite dimensional space into a finite dimensional optimization problem.

In regression in Sobolev spaces, we have demanded only smoothness constraint on regression function  $f \in \mathcal{F} = \left\{ f \in \mathcal{H}^m(\mathcal{Q}^q) : \|f\|_{Sob,m}^2 \leq L \right\}$ . Now, our estimators should satisfy additional constraints. We therefore focus on the imposition of additional constraint — isotonia — on nonparametric regression estimation and testing of this constraint. Two basic types of isotonia are monotonicity and concavity so we will concentrate mostly on them. We



would like to estimate subject to  $f \in \widetilde{\mathcal{F}} \subseteq \mathcal{F}$  where  $\widetilde{\mathcal{F}}$  combines smoothness with further functional properties and to test  $H_0 : f_0 \in \widetilde{\mathcal{F}}$ . We prove that balls of functions in Sobolev space are bounded and have bounded higher order derivatives. It permits us to estimate over rich set of functions with sufficiently low metric entropy and apply laws of large numbers and central limit theorems results.

## Reference

- [1] Yatchew A. *Semiparametric Regression for Applied Econometrician*. Cambridge University Press, 2003.
- [2] Bos L., Yatchew A. *Nonparametric least squares estimation and testing of economic models*. Journal of Quantitative Economics **13**, 81 – 131, 1997.

## L-MOMENTY A JEJICH NĚKTERÁ ZOBECNĚNÍ

Jan Pícek

TU Liberec, KAM, Hálkova 6, 461 17 Liberec

*E-mail:* jan.picek@vslib.cz

Momenty se tradičně užívají k charakterizaci rozdělení pravděpodobnosti. Koncept L-momentů pochází z různých výsledků o lineární kombinaci pořádkových statistik, které v roce 1990 sjednodil Hosking. Aplikace L-momentů se objevila od té doby v řadě oblastí, zejména v klimatologii, hydrologii a kontrole jakosti. Cílem tohoto příspěvku je ukázat možná zobecnění, která se v poslední době v literatuře objevila a která odstraňují některé nevýhody klasických L-momentů.

## Reference

- [1] Elamir E., Seheult A.H. *Trimmed L-moments*. Computational Statistics & Data Analysis **43**, 299 – 314, 2003.
- [2] Hosking J. *L-moments: Analysis and estimation of distributions using linear combinations of order statistics*. Journal of Royal Statistical Society **B 52**, 105 – 124, 1990.
- [3] Wang Q.J. *LH moments for statistical analysis of extreme events*. Water Resources Research **33**, 2841 – 2848, 1997.

## SIGNIFIKANTNOST REGRESNÍCH KOEFICIENTŮ URČENÝCH METODOU LWS

Plát Pavel

FJFI ČVUT, Trojanova 13, 120 00 Prague 2.

*E-mail:* plat@kmlinux.fjfi.cvut.cz

Pro odhad regresních koeficientů metodou nejmenších vážených čtverců ( $LWS$ ), který je zobecněním odhadu metodou nejmenších usekaných čtverců, byly již odvozeny (a publikovány) nejen základní asymptotické vlastnosti ( $\sqrt{n}$ -konzistence, asymptotická reprezentace), ale i další statistiky z oblasti diagnostiky regresního modelu (např. modifikovaná Durbin-Watsonova či Whiteova statistika, odhad rozptylu disturbance či odhad kovarianční matice). Tento příspěvek navazuje na dosud publikované výsledky o  $LWS$  a zabývá se metodami, které můžeme použít k otestování signifikantnosti jednotlivých koeficientů a celého modelu určeného pomocí  $LWS$ .

## Reference

- [1] Chatterjee S., Price B. *Regression Analysis by Example*. J. Wiley & Sons, New York, 1977.
- [2] Kalina, J. *Autocorrelated disturbances of robust regression*. Personal com., 2003.
- [3] Mašíček L. *Diagnostika a senzitivita robustních modelů*. PhD disertace, MFF UK, Praha, 2003.
- [4] Plát P. *Odhad metodou nejmenších vážených čtverců*. Diplomová práce, FJFI ČVUT, Praha, 2003.
- [5] Plát P. . *Modifikace Whiteova testu pro nejmenší vážené čtverce*. Sborník ROBUST 2004, J. Antoch a G. Dohnal ed., 291–298, 2004.
- [6] Rousseeuw P.J., Leroy A.M. *Robust Regression and Outlier Detection*. J.Wiley & Sons, New York, 1987.

## NĚKTERÉ PROBLÉMY DETEKCE ZMĚN PRO ZÁVISLÁ POZOROVÁNÍ

Zuzana Prášková

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

*E-mail*: Zuzana.Praskova@mff.cuni.cz

V příspěvku budou prezentovány některé asymptotické výsledky týkající se detekce změn v lokačním modelu, resp. v modelu lineární regrese, když chyby nejsou nezávislé, ale vykazují určitý typ slabé závislosti ( $m$ -závislost, případně  $\alpha$ -mixing). Bude porovnána asymptotická relativní eficeence odhadů bodu změny v  $L_1$  a  $L_2$  normě při nezávislých a závislých chybách pro různé typy rozdělení.

## APPROXIMATION OF STATIONARY DENSITY IN SOME AUTOREGRESSIVE MODELS

Pavel Ranocha

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

*E-mail*: ranocha@karlin.mff.cuni.cz

The contribution deals with methods of calculating of stationary density in some autoregressive models of time series. Exact solution of this problem is known in only few special cases. We propose algorithms which approximate such solution.

For AR(1) model we describe a method from Anděl, Hrach (2000) and show that under some conditions its speed of convergence is geometrical. We generalize this algorithm to AR(2) and multidimensional AR(1) models. Further we deal with the limit behavior of sequences  $\eta_1 + \rho\eta_2 + \rho^2\eta_3 + \dots + \rho^n\eta_{n+1}$ , where  $\eta_i$  are iid random variables and  $\rho \in (-1, 1)$ , see e.g. Haiman (1998). We show that the density of these partial sums converges to the density of stationary distribution of AR(1) process. In the final part we derive stationary marginal distribution in two nonlinear models – absolute autoregression AAR(1) and threshold model TAR(1) for some distributions of innovations.

## Reference

- [1] Anděl J., Hrach K. *On calculation of stationary density of autoregressive processes*. *Kybernetika* **36**, 311–319, 2000.
- [2] Haiman G. *Upper and lower bounds for the tail of the invariant distribution of some AR(1) processes*. *Asymptotic Methods in Probability and Statistics* **45**, B. Szyszkowicz ed., 723–730, Elsevier Science B.V., 1998.
- [3] Loges W. *The stationary marginal distribution of a threshold AR(1) process*. *Journal of Time Series Analysis* **25**, 103–125, 2004.
- [4] Tong H. *Non-linear Time Series. A Dynamical System Approach*. Clarendon Press, Oxford, 1990.

## MODEL VÝVOJE NEZAMĚSTNANOSTI V ČASE A OKRESECH

Soňa Reisnerová

ÚTIA AV ČR, Pod Vodárenskou Věží 4, 182 07 Praha 8

*E-mail*: sona.reisnerova@centrum.cz

Cílem příspěvku je zachytit časovou i prostorovou závislost souboru náhodných veličin. Mějme časové řady náhodných veličin - počtů událostí - v daných lokalitách. Počty modelujeme pomocí Poissonova rozdělení a zároveň chceme zachytit vazby mezi sousedními lokalitami. Pro intenzitu (parametr) Poissonova rozdělení volíme tvar obvyklý pro Coxův model, tj. funkce času násobena  $\exp(\text{funkce dalších vlivů})$ , kde v druhé části bude popsána uvažovaná vzájemná závislost lokalit, a to formou Markovova náhodného pole. Dostáváme tak jednoduchý model prostorového-časového procesu, zde s diskretním časem. Při řešení odhadu parametrů modelu nejprve uděláme odhad za situace, kdy předpokládáme nezávislost na okolí, tento odhad se dále pokusíme zpřesnit pomocí Metropolisova-Hastingsova algoritmu. Tato metoda

je demonstrována na reálných datech měsíčního vývoje přírůstků nezaměstnaných v České republice, které jsou po dobu několika let pozorovány ve všech okresech.

## ZMĚNY PARAMETRŮ WEIBULLOVA ROZDĚLENÍ

**Monika Rencová**

**Stav. Fak. ČVUT, kat. matematiky, Thákurova 7, 166 29 Praha 6**

*E-mail:* rencova@mat.fsv.cvut.cz

Pro modelování pravděpodobnostního chování meteorologických a hydrologických řad se často používá jiné rozdělení než normální, protože data vykazují výraznou šikmost. To platí například o maximálních a minimálních teplotách. Jedno z rozdělení, které můžeme použít je trojparametrické Weibullovo rozdělení.

Zajímáme-li se o změny v parametrech modelu, pak je výhodné sledovat jejich vývoj v čase například tak, že odhad provádíme v postupně se posouvajícím oknu. Ještě lepší odhad získáme, jestliže použijeme váhy podobně jako v jádrových odhadech hustot, viz Hall, Tajvidi (2000). Naším cílem bylo použít tyto metody pro teplotní řady, jejichž vlastnosti byly prezentovány na Robustu 04.

## PROBLÉMY HODNOCENÍ STRUKTURY NANOMATERIÁLŮ

**I. Saxl<sup>1,3,a</sup>, L. Ilucová<sup>1,b</sup>, V. Sklenička<sup>2,c</sup> a M. Svoboda<sup>2,d</sup>**

<sup>1</sup>Matematický ústav AV ČR, Žitná 25, 115 67 Praha,

<sup>2</sup>Ústav fyziky materiálů AV ČR, Žižkova 22, 616 62 Brno,

<sup>3</sup>MFF UK, KPMS, Sokolovská 83, 186 75 Praha 8 – Karlín

*E-mail:* <sup>a</sup>saxl@math.cas.cz, <sup>b</sup>ilucova@math.cas.z, <sup>c</sup>sklen@ipm.cz,

<sup>d</sup>svobm@ipm.cz

Materiály s velmi malým zrnem (řádově kolem  $1\mu\text{m}^3$  a menším - *nanomateriály*) jsou pro své výjimečné mechanické vlastnosti velmi perspektivní. Jejich složení i technologie je ovšem teprve ve stadiu výzkumu; jedním ze slibných postupů je metoda opakovaného protlačování (ECAP - equal channel angular pressing, protlačovací nástroj je zalomen do pravého úhlu, válcový polotovar je před opakovaným protlačováním otočen kolem své osy o určitý úhel). Struktura vzniklého zrna resp. subzrna, tj. jeho velikost, homogenita a prostorová orientace a následná stabilita této struktury (při ohřevu, zatěžování resp. obojím) jsou zřejmě zcela zásadní poznatky. Velmi široké možnosti poskytuje orientovaná elektronová mikroskopie, při níž je určována vzájemná

dezorientace  $\Delta$  subzrn a zrn. Jejich velikost a prostorová orientace jejich hranic tak může být měřena v širokém rozsahu hodnot  $\Delta$  ( $\Delta < 15^\circ$  odpovídá subzrnům,  $\Delta > 15^\circ$  zrnům). Kritickým problémem je nalezení optimálního počtu průchodů z hlediska stability struktury zrn a dobrých mechanických vlastností, zvláště za zvýšených teplot. Spoluprací dvou pracovišť AV ČR byla v uplynulých dvou letech provedena rozsáhlá studie struktury velmi čistého hliníku připraveného metodou ECAP za pokojové teploty s různým počtem průchodů, příp. dále zatěžovaného konstantním napětím při teplotě 473 K (vysokoteplotní creep). Metodika strukturního zkoumání využívala stereologické metody (paralelní měření ve třech význačných navzájem kolmých rovinách řezu a ve třech význačných navzájem kolmých směrech) a opírala se jednak o navržená schémata klasifikace teselací (tzv. *w-s diagramy*), jednak o rozsáhlou databázi prostorových teselací <http://fyzika.ft.utb.cz/voronoj/> vytvořenou v uplynulých letech. S jejich pomocí bylo možné získat dobré odhady prostorové velikosti subzrn i zrn i představu o jejich tvarech a o prostorové orientaci jejich hranic. Zásadním problémem je extrémní nehomogenita struktury i po vyšším počtu průchodů; její vliv na mechanické vlastnosti je podle dosavadních výsledků nejasný, nicméně naznačuje, že dokonalá homogenita by nemusela být optimální.

## SOFTWARE PRO ODHADY ROC KŘIVEK A AUC

Hana Skalská

FIM UHK, Rokitanského 62, 500 03 Hradec Králové

*E-mail:* hana.skalska@uhk.cz

Diskriminační analýza patří mezi úlohy klasifikace objektů, která řeší zařazení prvku na základě vektoru pozorování do jedné ze tříd, které jsou předem známy. Statistické i nestatistické metody diskriminace prvků zpravidla umožňují sestavit diskriminační model který minimalizuje očekávanou celkovou ztrátu, plynoucí z nesprávného zařazení prvků. Pro optimální diskriminační pravidlo je třeba znát apriorní pravděpodobnosti skupin mezi kterými se diskriminace provádí a ceny, spojené s jednotlivými chybami klasifikace. Při změně těchto vstupních parametrů není zaručeno, že sestavený model zůstává optimální i pro novou situaci. Jako míry kvality diskriminačního modelu a jeho prediktivních vlastností se doporučují různé charakteristiky. Kvantitativní míry lze doplnit vizuálně pomocí ROC křivky. ROC má praktické využití zejména při diskriminaci mezi dvěma skupinami. Popisuje (zobrazuje) vlastnosti klasifikačního modelu na celé množině rozhodovacích možností. Tento příspěvek se zabývá popisem základních vlastností ROC křivky, dualitou mezi ROC křivkou a očekávanými náklady a podává přehled softwarových mož-

ností při výpočtech křivky ROC a plochy pod křivkou, AUC. Jsou zmíněny možnosti stanovení ROC křivky pro různé typy klasifikačních algoritmů a možnost jejich využití pro sestavení tak zvaného hybridního klasifikačního modelu. Je sestaven přehled softwarových produktů pro různé typy odhadů AUC a je provedeno porovnání jejich možností.

## STOCHASTICKÁ VERZE KLASICKÉHO MODELU VÝVOJE EPIDEMIE

**Jakub Staněk**

**MF F UK, KPMS, Sokolovská 83, 186 75 Praha 8 – Karlín**

*E-mail:* stanekj@mff.cuni.cz

Tento příspěvek se zabývá stochastickou verzí klasického modelu vývoje epidemie. Tato verze je dána 3-dimenzionální stochastickou diferenciální rovnicí

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

kde  $X_t$  je počet zdravých jedinců, u kterých hrozí nákaza,  $Y_t$  je počet infikovaných jedinců,  $Z_t$  je počet imunních jedinců a pro velikost populace  $N_t = X_t + Y_t + Z_t$  požadujeme, aby  $a \leq N_t \leq b$  pro  $t \geq 0$  a  $0 \leq a \leq n_0 \leq b < \infty$ .

Předpokládáme-li, že koeficienty  $\beta$  a  $\sigma$  jsou v jistém smyslu Lipschitzovské a omezené, pak lze dokázat existenci a jednoznačnost silného řešení výše uvedené soustavy. Uvedené podmínky jsou shrnuty do věty, která umožňuje následující volby koeficientu  $\beta$ :

$$\beta \in \mathbb{R}^+, \beta(x), \beta(y), \beta(z) \in \mathcal{C}(\mathbb{R}, \mathbb{R}^+), \beta(x, y, z) \in \mathcal{C}(\mathbb{R}^3, \mathbb{R}^+), \beta(x, y, z, t) \in \mathcal{C}(\mathbb{R}^3 \times \mathbb{R}^+, \mathbb{R}^+) \text{ a } \beta(x, y, z, t) \text{ je progresivní funkcionál na } \mathcal{C}(\mathbb{R}^3, \mathbb{R}^+) \times \mathbb{R}^+.$$

## KLASIFIKAČNÍ STROMY VE SPOLEHLIVOSTI SOFTWARE

**Jan A. Strouhal**

**ČVUT, FS ÚTM, Karlovo nám. 13, 120 00 Praha 2**

*E-mail:* jstrouhal@email.cz

Příspěvek se zabývá konstrukcí klasifikačních stromů pro použití ve spolehlivosti software. V první části je popsána základní konstrukce, která počítá s expertním počtem i hodnotami hranic metrik, v další části jsou popsány vylepšení této základní konstrukce vyhledáním optimálních hranic metrik. Příspěvek také popisuje rozšíření základní klasifikační škály na více skupin.

**FACTORS INFLUENCING ENTERPRISES'  
STANDING ON THE MARKET****Pavel Stríž, Jitka Kloudová, Kamil Dobeš****ÚIS FAME UTB, Mostní 5139, 760 01 Zlín***E-mail: striz@fame.utb.cz*

A market research investigation was conducted in 2003 to discover which factors are the most significant for maintaining one's pace (and place) with the competition in 276 Czech enterprises. The aim of this research study was to map the environment of Czech enterprises from the perspective of the situational standing of that enterprise on the Czech market. Using the GE Matrix Model, it investigated whether the set of professionally-related activities in which the enterprise operates has an influence on the situational standing of these enterprises on their markets, i.e. the quality of the enterprise's management, the level (and quality) of marketing in the enterprise, the exploitation of the Internet in the marketing activities of that enterprise, and at the same time, control checks were made of the marketing activities through the performance of a Marketing Audit.

**GENE EXPRESSION DATA ANALYSIS FOR IN VITRO TOXICOLOGY****Petr Šimeček****ÚTIA AV ČR, Pod Vodárenskou Věží 4, 182 07 Praha 8***E-mail: simecek@karlin.mff.cuni.cz*

Práce demonstruje použití analýzy DNA–mikropolí k určení genů, jejichž exprese je ovlivněna alergeny. Dendritické buňky (CD34–DC) získané z pupečnickové krve 20 jedinců byly vystaveny po jeden z šesti různých časových intervalů (od 30 minut do 24 hodin) působení jedné z šesti chemikálií. Následně byla z každého vzorku extrahována mRNA a analyzována pomocí cDNA–mikropolí (11395 genů). Cílem bylo zjistit exprese kterých genů je prokazatelně ovlivněna působením čtyř z šesti chemikálií (alergenů) a zároveň netknuta ostatními dvěmi substancemi. Diskutována je normalizace polí, stabilizace rozptylu a detekce měření se zvýšeným šumem.

**STATISTICAL ANALYSIS OF COMPULSIVE CHECKING  
BEHAVIOR IN RODENTS – FRAILTY MODELS****Marie Šimečková****MFF UK, KPMS, Sokolovská 83, 186 75 Praha 8 – Karlín***E-mail: simecko@karlin.mff.cuni.cz*

Opakovaná měření, vážící se k jednomu objektu, se ve statistické praxi objevují velmi často. V případě, že výsledky měření jsou cenzorovaná, lze k jejich modelování použít tzv. frailty model - Coxův regresní model s náhodnými efekty.

V příspěvku jsou frailty modely aplikována na reálná data z biologie. Do pokusu bylo zahrnuto 22 krys, u poloviny z nich byl pomocí chemikálie quinpirol navozen stav odpovídající psychické nemoci obsedantně-nutková porucha (OCD - obsessive-compulsive disorder). Cílem je zjistit vliv quinpirolu na proces učení a prostorovou paměť hlodavců. Každé zvíře bylo během čtyř dnů 24-krát ponořeno do Morrisova vodního bludiště a bylo měřeno, za jak dlouho nalezne pod vodou skrytou plošinu. Protože po 90 sekundách byly krysy, které plošinu nenalezly, z vody vyjmuty, data jsou zprava cenzorovaná. Bylo zjištěno, že quinpirol nemá významný vliv na schopnost učení krys, ale způsobuje zapomínání během delších přestávek mezi pokusy.

## Reference

- [1] Therneau T. M., Grambsch P. M. *Modeling Survival Data: Extending the Cox Model*. Springer-Verlag, New York, 2000.
- [2] Gray R. J. *Flexible methods for analyzing survival data using splines, with applications to breast cancer prognosis*. J. Amer. Stat. Assoc. **87**, 942–951, 1992.

## DIDAKTIKA PRAVDĚPODOBNOTI A STATISTIKY

Milena Špínková

MÚ AV ČR, Žitná 25, 115 67 Praha 1

*E-mail*: Milena.Spinkovalucova@vsfs.cz

Množství dat statistického charakteru produkované medií, reklamními agenturami i neziskovými organizacemi a rostoucí složitost běžného denního života klade značné nároky na *statistické myšlení* každého jednotlivce (slovo *statistické* je zde zkratkou za pravděpodobnostní a statistické). To sestává jednak z vědomí variability v průběhu každé naší činnosti a posouzení jejích příčin, jednak o alespoň elementární znalosti metod sběru, úpravy, interpretace a prezentace dat. Naše současná rodinná i školní výchova však vytvoření tohoto způsobu myšlení téměř vůbec nenapomáhá, neboť je striktně deduktivní, založená na jednoznačné vazbě mezi příčinami a následky. Výuce pravděpodobnosti a statistiky je na základních školách věnována zcela minimální pozornost navzdory tomu, že náš život je téměř spojitým řetězcem nejistých jevů s často nenapravitelnými důsledky. Zlepšení této situace je dlouhodobý proces, vyžadující především výchovu učitelů. Výchozím bodem by ovšem měla být skutečnost, že o nejistých jevech toho děti z vlastní zkušenosti od



nejútlejšího věku hodně vědí a škola i rodina na tom svou výchovu ke statistickému myšlení mohou založit. Je dobré si připomenout výrok Galilea Galilei: *Není možné člověka něco naučit, lze mu pouze pomoci objevit znalosti v něm samém.*

## EVOLUČNÍ ALGORITMY

Josef Tvrdík

Ostravská universita, Přír. fak., kat. informatiky, 30. dubna 22,  
701 03 Ostrava

*E-mail:* tvrdik@osu.cz

V příspěvku budou vysvětleny základní myšlenky evolučních algoritmů jako účinného heuristického přístupu k hledání globálního extrému [1], [2], [5]. Dále budou popsány nejdůležitější třídy evolučních algoritmů, a to genetické algoritmy [3], evoluční strategie [4], diferenciální evoluce [6], samoorganizující migrační algoritmus [9], [10] a adaptivní algoritmy využívající soutěž a spolupráci různých heuristik [7], [8]. Uvedeny budou i výsledky těchto algoritmů v odhadu parametrů nelineárních regresních modelů.

## Reference

- [1] Bäck T. *Evolutionary Algorithms in Theory and Practice*. Oxford University Press, New York, 1996.
- [2] Goldberg D.E. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Reading, Addison Wesley, 1989.
- [3] Holland J. H. *Adaptation in Natural and Artificial Systems*. The University of Michigan Press, Ann Arbor, MI, 1975.
- [4] Kvasnička V., Pospíchal J., Tiňo P. *Evolučné algoritmy*. STU, Bratislava, 2000.
- [5] Michalewicz Z. *Genetic Algorithms + Data Structures = Evolution Programs*. Berlin, Springer Verlag, 1992.
- [6] Storn R., Price K. *Differential evolution – a simple and efficient heuristic for global optimization*. J. Global Optimization **11**, 341–359, 1997.
- [7] Tvrdík J., Mišík L., Křivý I. *Competing heuristics in evolutionary algorithms*. In: Sinčák P. et al. (Eds.), 2nd Euro-ISCI Intelligent Technologies – Theory and Applications, IOS Press, Amsterdam, 159–165, 2002.
- [8] Tvrdík J. *Competition and cooperation in evolutionary algorithms: A comparative study*. In: Proceedings of MENDEL 2005, 11-th Int. Conference on Soft Computing (Matoušek R. and Ošmera P. eds), Technical University Press, Brno, 108–113, 2005.

- [9] Zelinka I., Lampinen J. *SOMA – self-organizing migrating algorithm*, In: Proceedings of MENDEL 2000, 6-th Int. Conference on Soft Computing, 177–187. Brno: Technical University, 2000.
- [10] Zelinka I. *Umělá inteligence v problémech globální optimalizace*. BEN, Praha, 2002.

## **SIMULAČNÍ STUDIE ROBUSTNOSTI ODHADŮ PRAHOVÝCH PARAMETRŮ NĚKTERÝCH ROZDĚLENÍ**

**Vladimír Václavík**

**ZČU, kat. matematiky, POBox 314, 306 14 Plzeň 1**

*E-mail:* bos@kma.zcu.cz

Simulační studie ukazují, že pro výběry běžného rozsahu ( $n \leq 100$ ) je obtížné odlišit logaritmicko-normální, Weibullovo a gama rozdělení. Uvažujeme tříparametrické varianty těchto rozdělení s prahovým parametrem. Zkoumáme různé bodové a intervalové odhady prahových parametrů (a obecněji též 100p% kvantilů pro velmi nízká  $p$ ) pro tato rozdělení a robustnost těchto odhadů vůči chybnému výběru modelu. Zabýváme se též případy, kdy výběr pochází ze zobecněného logaritmicko-normálního nebo zobecněného gama rozdělení nebo je směsí výběrů z rozdělení s různými parametry.

## **THE PROSPECTOR – BEYOND CLASSIFICATION**

**Pavel Vaněček**

**MFF UK, KPMS, Sokolovská 83, 186 75 Praha 8 – Karlín**

*E-mail:* vanecek@karlin.mff.cuni.cz

Contemporary computational statistics “suffers” from an abundance of data. This surplus requires completely new approach to data mining: Not information but knowledge is what matters. The aim of modern data mining is not only searching for patterns but also presenting them in some comprehensible form. The typical receiver of the results has little statistical background but he or she has an experience and an intuition so the results should be delivered in his or her “language”.

This presentation belongs to the data mining field. In cooperation with the development department of market research company TNS AISA, we develop a tool for data description - The Prospector. This software package enables us to find and identify important patterns in data. The concept is simple: We are not interested in the description of all items but we focus on some well defined subsets only.

The Prospector stands on the border between clustering and classification: We are searching for clusters (called “nuggets”) and instead of similarity of members of one nugget we require their superiority towards a target variable. The key objective in this algorithm is scaling of the input variables. Since the clustering part of The Prospector allows only categorical variables due to computational efficiency, the categorization and ordering is crucial. We employ a genetic algorithm which ensures fast and efficient way of solving the underlying optimization problem. Afterwards, these inputs are searched for the subgroups that are relevant to a target variable. Found subgroups are examined and sorted via some desired criterion which brings up traditional trade-off between size of a particular subgroup and its accuracy (or similarity, purity). We would like to present both statistical and computational aspects of The Prospector, show its potential for solving practical problems, and perform a case study on real business data.

### CONSISTENCY OF EMPIRICAL DISTRIBUTION FUNCTION OF RESIDUALS IN LINEAR REGRESSION

Jan Ámos Víšek

Fac. of Social Sciences, Charles University, Smetanovo nábřeží 6,  
110 01 Prague 1

Inst. of Information Theory and Automation, Academy of Sciences  
of the Czech Republic, Pod Vodárenskou Věží 4, 182 07 Praha 8

*E-mail:* visek@mbox.fsv.cuni.cz

Let  $N$  denote the set of all positive integers,  $R$  the real line and  $R^p$  the  $p$ -dimensional Euclidean space. The linear regression model given as

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

will be considered. We shall assume that:

**C1** *The sequence  $\{(X_i', e_i)'\}_{i=1}^\infty \subset R^{p+1}$  is sequence of independent and identically distributed random variables.*

Notice please that our notation covers both the model with intercept as well as without it. In the former case, i.e. when we consider the model with intercept, we have to assume that the first coordinate of explanatory variables  $X_i$  is degenerated and equal to 1. The Condition **C1** doesn't assume the orthogonality between  $X_i$ 's and  $e_i$ 's, i. e.  $\mathbb{E} \{X_1 \cdot e_1\}$  need not be zero. It allows to employ the result of paper also for *instrumental variables* and their robust version.

For any  $\beta \in R^p$   $r_i(\beta) = Y_i - X_i'\beta$  denotes the  $i$ -th residual. Then  $\sqrt{n}$ -consistency of the empirical distribution function of residuals in linear regression model (1) is proved. The convergence is uniform with respect to the regression coefficients and to the argument of d.f. An example which was the inspiration for deriving the result will be given.

## SHLUKOVÁNÍ ČASOVÝCH ŘAD POISSONOVÝCH NÁHODNÝCH VELIČIN

Petr Volf

ÚTIA AV ČR, Pod Vodárenskou Věží 4, 182 07 Praha 8

*E-mail:* volf@utia.cas.cz

V příspěvku na Robustu 2004 jsem řešil otázku, jak vybrat optimální počet komponent modelu, a to speciálně v modelech směsí a tedy i shlukové analýzy. I když jsem se jako obvykle nevyhnul chybám v textu, došel jsem k snad rozumnému závěru, že kriteria typu BIC, Bayesův faktor a pod. jsou užitečná, aspoň pro základní orientaci (zde již není zaručena jejich konzistence), Reversible Jump MCMC (P. Green et al) jsou teoreticky v pořádku, ale skrývají v sobě jednak příliš volnosti pro volbu apriorních vstupů a jsou i příliš výpočetně náročná (což není problém času, ale sledování toho, co se vlastně děje během výpočtu), takže nejlépe je volit střední cestu, tj. snažit se maximalizovat aposteriorní pravděpodobnost (odtud vycházejí aproximace vedoucí k prvně zmíněným kritériím a sem směřuje nakonec MCMC počítání).

Tuto ideu je snadné předvést na jednoduchém příkladu K-means shlukování, ale už hůře v případě shlukování náhodných posloupností (ať už z důvodu vhodného výběru priorů nebo i řízení optimalizace: "Simulované žihání je umění samo o sobě" - volně dle M. Janžury, Robust 1990). Pro ukázkou budou použita stejná data jako analyzuje S. Reisnerová, a to vývoj přírůstků počtu nezaměstnaných v různých okresech ČR modelovaný jako řady Poissonových náhodných veličin. Cílem je najít okresy s podobným vývojem. Zároveň lze ještě uvnitř každého shluku definovat tzv. heterogenitu jako multiplikativní parametr, který pro řady (tj. okresy) s podobným trendem charakterizuje jejich měřítko.

## NIEKTORÉ MATEMATICKO-ŠTATISTICKÉ MODELY KALIBRÁCIE

Gejza Wimmer

PřF MU, KAM, Janáčkovo nám. 2a, 662 95 Brno

*E-mail:* wimmer@math.muni.cz

Kalibrácia je komplex veľmi vážnych technických, legislatívnych, metrologických, ekonomických, ale aj matematicko-štatistických problémov. Tieto dodnes nie sú uspokojivo vyriešené. Práve matematicko-štatistické postupy môžu priniesť pokrok pri riešení kalibračných problémov, ktorých ťažiskom je získanie kalibračnej krivky (jej vhodného odhadu) a vyhodnotenie meraní realizovaných kalibrovaným meradlom. Prezentovať sa bude matematicko-štatistické modelovanie kalibračného problému a niektoré výsledky odhadov parametrov kalibračnej krivky, resp. návrh matematicko-štatistických postupov na vyhodnotenie meraní pomocou kalibrovaného prístroja.

## **ANALÝZA ROZPTYLU A ZMIEŠANÝ LINEÁRNY MODEL**

**Viktor Witkovský**

**ÚM SAV, Dúbravská cesta 9, 841 04 Bratislava**

*E-mail:* witkovsky@savba.sk

Zmiešaný lineárny model je zovšeobecnením klasického lineárneho regresného modelu, ktoré umožňuje zaviesť do modelu bohatšiu kovariančnú štruktúru, ktorá takto lepšie modeluje závislosť medzi uvažovanými náhodnými premennými. Z iného hľadiska možno zmiešaný lineárny model chápať ako zovšeobecnenie modelov analýzy rozptylu.

Vo svojej všeobecnosti prináša takýto model aj nové problémy v oblasti odhadovania parametrov a testovania hypotéz o parametroch. Exaktné riešenia pre malé rozsahy výberov sú známe len v špeciálnych situáciách. V prednáške spomenieme prehľad základných metód odhadovania parametrov prvého a druhého rádu (variančných komponentov), ako aj metódy testovania hypotéz o týchto parametroch modelu.

Uvedieme niekoľko príkladov, ktoré prirodzene vedú k zmiešanému lineárnemu modelu. Typickou oblasťou aplikácii tohto modelu sú napr. poľnohospodárske experimenty, avšak uplatnenie zmiešaného lineárneho modelu možno ilustrovať aj v takých oblastiach ako je analýza výsledkov experimentálnych meraní pri určovaní rýchlosti svetla.

## **Reference**

- [1] MacKay R.J., Oldford R.W. *Scientific method, statistical method and the speed of the light*. Statistical Science **15**, 254–278, 2000.
- [2] Pinheiro J.C., Bates D.M. *Mixed-Effects Models in S and S-Plus*. Springer-Verlag, New York, 2000.
- [3] Rao C.R., Kleffe J. *Estimation of Variance Components and Its Applications*. North-Holland, Amsterdam, 1988.
- [4] Searle S.R., Casella G., McCulloch C.E. *Variance Components*. Wiley, New York, 1992.

## Abstrakty dodané po termínu

### ZMĚNY DENNÍCH MAXIM A MINIM TEPLoty VZDUCHU

Milena Kovářová

UFB JU, Zámek 136, 373 33 Nové Hrady

*E-mail:* kovarova@greentech.cz

Příspěvek porovnává maximální a minimální teploty vzduchu naměřené na Mokřých Loukách s hodnotami naměřenými na několika jiných meteorologických stanicích. Ukazuje, že změna klimatu se neprojevuje všude stejně, ale závisí na místních podmínkách dané stanice.

### MĚŘENÍ MULTIKOLINEARITY S POMOCÍ DETERMINANTŮ KORELAČNÍ MATICE

Jiří Mihola

Vysoká škola ekonomie a managementu, Centrum ekonomických studií, I.P.Pavlova 3, 120 00 Praha 2

*E-mail:* Jiri.Mihola@quick.cz

Přednáška obsahuje odvození postupu na základě ortogonální úlohy, porovnání s jinými způsoby měření, vazbu na analýzu rozptylu, zobrazování výsledků pro  $n$ -rozměrný prostor, konkrétní příklady použití apod.

### KŘIVKY TYPU ROC JAKO NÁSTROJ PRO MĚŘENÍ A POROVNÁNÍ KVALITY NA JEVECH KOMPUTAČNÍ LINGVISTIKY

Pavel Pecina a Pavel Schlesinger

Ústav formální a aplikované lingvistiky, Univerzita Karlova

*E-mail:* pavel.schlesinger@mff.cuni.cz

Jedním z témat komputační lingvistiky je snaha o automatickou detekci a popis lingvistických jevů. Například nás může zajímat *kolokace*, tj. zda se dvojice či delší kombinace slov spolu vyskytuje v jazyce pravidelně a zda takové ustálené sousloví mění význam.

Matematickým modelem jevu je tzv. *asociační míra*, která číselnou hodnotou udává odhad síly kolokace. V současné době jsou známé již desítky různých asociačních měr, z čehož vyplývá snaha o porovnání jejich výkonu při hledání kolokace.

Cílem příspěvku bude ukázat spojitost dané úlohy s klasifikační úlohou a možnosti, které pro řešení podobných úloh může nabídnout aplikace křivky typu ROC. Při empirickém ověřování vycházíme z experimentu na ručně anotovaných textových datech Pražského závislostního korpusu, verzi 2.0.