

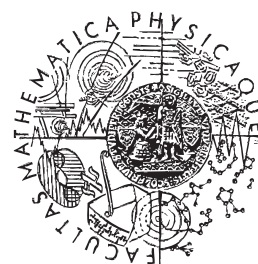
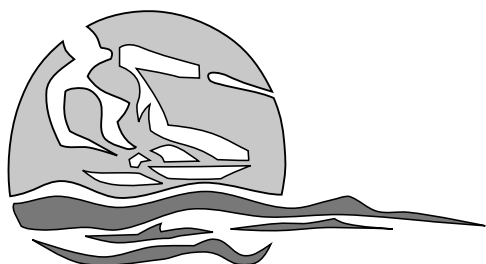
ROBUST 2012

Obec Němčický leží v členitém terénu Kněžského vrchu. Připomíná se roku 1349 a následujícího roku je v pramenech uváděna viniční hora Skalica. O kvalitě zdejších pěstitelských poloh svědčí, že vinice tu mají jak brněnský měšťan, tak samotné město Brno a Kyjov, bolearadický kostel, brněnský jezuité a další instituce. O zlepšení vinařské kultury se v 16. století přičinili novokřtění. 166 ha vinic, o něž pečuje 257 vinařů je rozloženo v těchto tratích:

1. Dlouhé kněžské - Malé odměry
2. Kněžské za humny
3. Hrušový - Kocůrky
4. Puclety
5. Nad Ochoze
6. Za louky
7. Odměry
8. Staré hory
9. Krátké kněžské - Novosady nad presůzy
10. Nové hory
11. Nad Ochůzky
12. Fílliberky
13. Šumberky
14. Kolberk
15. Zbavce - Novosady
16. Skalice příční
17. Skalice u věřáka
18. Tálečky pod Šumperky - Dlouhé táčky
19. Růženy - Novosadky
20. Špigle - Bočky
21. Soudná
22. Veselý

Místní pěstitelé uchovávali víno ve svých domech, velmi často společně s jinými plodinami. Jen ti největší vinaři si na okrajích dědiny "budovali" sklepy s prostornými lisovnicami a s patrovými pres-hausy, jejichž stroje účelově řešené průřelí je ozdobeno židram, které chrání vchod do lisovny. Sklepy jsou kopány v pískovcové podloží a mohou být celé nebo zčásti vyzděny. Zvláštností některých zdejších lisoven jsou boční, pouze zvenku přístupná schodiště na sýpku v patře, zabudovaná do pláště stavby.

Literatura: J. Matuzzková a V. Kovářů, Vinohradnické stavby. ERA, Brno, 2004.



ROBUST 2012 - PROGRAM				
NEDĚLE	DOPOLEDNE			
	10.00 - 12.00		registrace	
oběd	12.00 - 13.15			
NEDĚLE	ODPOLEDNE			
	13.30 - 13.45	15	J. Kyselý a J. Pícek	Zahájení - Extreme value analysis in climatology
	J. KYSELÝ			Extremes : From regional analysis to the peaks-over-threshold method
	13.45 - 14.15	30	M. Hanel	Regional block-maxima modelling of precipitation extremes in climate model simulation
	14.15 - 15.00	45	M. Roth	Regional peaks-over-threshold model in non-stationary climate
	15.00 - 15.30	30	M. Schindler	How to choose threshold in a POT model?
káva	15.30 - 15.50			
	J. PÍCEK			Extremes : Covariates
	15.50 - 16.35	45	S. Begueria	Covariate-dependent modelling of extreme events by nonstationary POT analysis
	16.35 - 17.20	45	P. Jonathan	Modelling covariate effects in extremes
	17.20 - 17.50	30	J. Dienstbier	Covariate effects in extremes - remarks and theory
	17.50 - 18.20	30		Discussion
večeře	18.30 - 19.30			
NEDĚLE	VEČER			
	M. HANEL			
	20.00-22.00	120	L. Metelka	Změna klimatu - mýty, fakta, statistika

PONDĚLÍ	DOPOLEDNE			
	J. CHAJDIÁK			
	9.00 - 9.15	15	J. Antoch	ROBUST 2012 : Zahájení
	9.15 - 9.45	30	K. Hron	Předzpracování kompozičních dat
	9.45 - 10.15	30	E. Fišerová	Ortogonalní regrese pro 3-složkové kompoziční data využitím lineárních modelů
přestávka	10.15 - 10.20			
	K. HRON			
	10.20 - 10.30	10	A. Kalivodová	Užití kompozičního biplotu při analýze medicínských dat (poster)
	10.30 - 10.40	10	K. Hružová	Bilance a bilanční dendrogram kompozičních dat (poster)
	10.40 - 10.50	10	P. Kynčlová	Dirichletovo rozdělení vzhledem k Aitchisonově míře na simplexu (poster)
	10.50 - 11.00	10	S. Donevska	Calibration between log-ratios of parts of compositional data (poster)
káva	11.00 - 11.20			
	Z. FABIÁN			
	11.20 - 11.50	30	J. Klaschka	Podruhé o výpočtu Blakerova konfidenčního intervalu: Balíček BlakerCI a jiné resty
	11.50 - 12.20	30	J. Běláček	O vizualizaci statistických dat
oběd	12.30 - 13.45			
PONDĚLÍ	ODPOLEDNE			
	J. ANTOCH			
	14.00 - 15.00	60	M. Hladík a M. Černý	Algoritmy, složitost a intervalová data
přestávka	15.00 - 15.10			
	M. FRIEŠL			
	15.10 - 15.40	30	L. Fajfrová	Náhodná procházka na hierarchické grupě
	15.40 - 16.10	30	M. Grendár	Prevádzať p-hodnotu na Bayesov faktor?
káva	16.10 - 16.30			
	M. GRENDÁR			
	16.30 - 17.00	30	R. Rosipal	Multi-way data analysis for advanced physiological estimation of cognitive status
	17.00 - 17.30	30	Z. Pawlas	Odhad rozdělení latence odezvy neuronu
večeře	18.00 - 19.00		večeře	
PONDĚLÍ	VEČER			
	20.00 - ??.??		J. Stávek	Řízená ochutnávka místních vín

ÚTERÝ	DOPOLEDNE			
	V. WITKOVSKÝ			
	8.30 - 9.00	30	V. Beneš	Náhodné množiny s doprovodnými proměnnými a redukce dimenze
	9.00 - 9.30	30	M. Prokešová	Statistická inference pro Coxovy bodové procesy s Lévyho bází
	9.30 - 10.00	30	J. Dvořák	Časoprostorové Coxovy bodové procesy s Lévyho bází
káva	10.00 - 10.20			
	R. ROSIPAL			
	10.20 - 10.50	30	T. Lechner	Analýza časových řad formální komunikace obcí
	10.50 - 11.20	30	Š. Hudecová	Testování změn v binárních autoregresních modelech
	11.20 - 11.50	30	D. Jarušková	Detekce změny kovariančního operátoru
oběd	12.00 - 13.15		oběd	
ÚTERÝ	ODPOLEDNE			
	D. JARUŠKOVÁ			
	14.00 - 14.30	30	M. Hušková	Sekvenční testování stability ve funkcionálním modelu CAPM
	14.30 - 15.00	30	Z. Prášková	Robustní monitorování stability modelu CAPM
	15.00 - 15.30	30	J. Antoch	O segmentaci velmi dlouhých časových řad
káva	15.30 - 15.50			
	M. HUŠKOVÁ			
	15.50 - 16.00	10	H. Horáková	Detekce změny ročního chodu průtokových řad (poster)
	16.00 - 16.10	10	K. Starinská	Parameters estimates for change-point detection problem in AR series (poster)
	16.10 - 16.20	10	P. Laník	Semimetrický přístup k odhadování časových radov (poster)
	16.20 - 16.30	10	R. Sabolová	Testy pre regresné kvantily založené na metóde sedlového bodu (poster)
	16.30 - 16.40	10	P. Novák	Regrese v modelech oprav (poster)
	16.40 - 16.50	10	M. Tučková	Design experimentu pro regresní modely s podmínkami typu I (poster)
přestávka	16.50 - 17.00			
	V. BENEŠ			
	17.00 - 17.10	10	S. Bělašková	Optimal site for cardiac pacing in children (poster)
	17.10 - 17.20	10	K. Fačevicová	Použití logistické regrese pro diagnostiku výskytu rakoviny prostaty (poster)
	17.20 - 17.30	10	D. Mlčůchová	Aplikované statistické metody v analýzách onkologických dat zvířecích experimentů (poster)
	17.30 - 17.40	10	V. Sečkárová	Dynamic bayesian estimation in diffusion networks (poster)
	17.40 - 17.50	10	D. Stibůrek	Asymptotická ekvivalence statistik spojených difúzních procesů (poster)
	17.50 - 18.00	10	M. Zikmundová	Užití částicového Metropolisova Hastingsova algoritmu ve stoch. geometrii (poster)
	18.00 - 18.10	10	B. Shokirov	A lower bound for the mixture parameter and its estimator (poster)
večeře	18.15 - 19.15			
ÚTERÝ	VEČER			
	G. DOHNAL			
	20.00 - 20.45	45	P. Schlesinger	SAS
	20.45 - 21.30	45	T. Jurczyk	STATISTICA

STŘEDA	DOPOLEDNE			
	B. ŠEDIVÁ			
	8.30 - 9.30	60	D. Hlubinka, S. Nagy a O. Vencálek	O hloubce dat (zobecněná poloprostorová hloubka; konvergence; klasifikace)
káva	9.30 - 9.50			
	T. CIPRA			
	9.50 - 11.05	75	R. Mesiar	Kopule ako nástroj modelovania štruktúry stochastickej závislosti náhodných vektorov
přestávka	11.05 - 11.10			
	11.10 - 12.10	60	M. Omelka	Kopule, parciální a podmíněné korelační koeficienty
oběd	12.15 - 13.00			
STŘEDA	ODPOLEDNE			
výlet	13.00 - 19.30		výlet	
STŘEDA	VEČER			
večeře	20.00 - 22.00			

ČTVRTEK	DOPOLEDNE			
	A. KOMÁREK			
	8.30 - 9.00	30	M. Friesl	Testování normality ze zaokrouhlených dat
	9.00 - 9.30	30	M. Kulich	Jednovýběrový vážený t-test pro pozorování s různými rozptyly
	9.30 - 10.00	30	M. Chvosteková	Simultánne testovanie strednej hodnoty a variancie normálneho rozdelenia
káva	10.00 - 10.20			
	M. KULICH			
	10.20 - 10.50	30	V. Witkovský	Exaktné testy a konfidenčné oblasti pre parametre normálneho lineárneho modelu
	10.50 - 11.20	30	Z. Fabián	Resuscitace momentové metody
	11.20 - 11.50	30	P. Lachout	Lineární regrese trochu jinak
oběd	12.00 - 13.15			
ČTVRTEK	ODPOLEDNE			
	P. LACHOUT			
	14.00 - 15.00	60	T. Cipra	Některé kvantitativní aspekty penzí
přestávka	15.00 - 15.05			
	J. LUHA			
	15.05 - 15.35	30	M. Pešta	Asymptotic consistency and inconsistency of the chain ladder
	15.35 - 16.05	30	B. Šedivá	Stabilita optimální volby portfolia Markowitzova modelu
	16.05 - 16.35	30	N. Kasperíková	Některé prostředky pro analýzu sekvencí
káva	16.35 - 16.55			
	J. KLASCHKA			
	16.55 - 17.20	25	E. Michalíková	The factors of growth of small family businesses
	17.20 - 17.45	25	M. Žambochová	Modifikace algoritmu FEKM
	17.45 - 18.10	25	S. Koróny	Problematika riešenia efektívnosti verejných vysokých škôl na Slovensku
	18.10 - 18.25	15	J. Chajdiak	Nezdanený objem hrubého domáceho produktu SR a DPH v letech 1996-2011
ČTVRTEK	VEČER			
večeře	20.00 - ??.??		J. Wimmer a kol.	Závěrečná večeře

PÁTEK	DOPOLEDNE			
	M. OMELKA			
	9.00 - 9.30	30	I. Kasanický	Identifikace netypického chování fotovoltaických elektráren
	9.30 - 10.00	30	O. Konár	Optimalizace osazování odběrných míst inteligentními plynoměry
	10.00 - 10.30	30	G. Dohnal	Adaptivní regulační diagramy
	10.30 - 10.40	10	J. Král	Metodika komplexního návrhu regulačního diagramu (poster)
	10.40 - 10.50	10	E. Cézová	Ekonomicko-statistická optimalizace regulačních diagramů (poster)
káva	10.50 - 11.10			
	G. DOHNAL			
	11.10 - 11.40	30	P. Tuček	Optimální návrh měření sigmoidálních funkcí
	11.40 - 12.10	30	D. Legát	Statistická analýza obrazu a kontrola jakosti
	12.10 - 12.40	30	P. Marek	Modelování a predikce výsledků hokejových zápasů
	12.40 - 12.45	5	J. Antoch	Ukončení
oběd	13.00 - 14.00			

ROBUST 2012

Sborník abstraktů

Antoch Jaromír	
<i>O segmentaci velmi dlouhých časových řad</i>	4
Beguiria Portugués Santiago	
<i>Covariate-dependent modelling of extreme events by nonstationary POT analysis</i>	4
Běláček Jaromír	
<i>O vizualizaci statistických dat</i>	4
Bělašková Sylvie, Janoušek Jan, Volaufová Júlia	
<i>Optimal site for cardiac pacing in children</i>	5
Beneš Viktor, Šedivý Ondřej, Staněk Jakub	
<i>Náhodné množiny s doprovodnými proměnnými a redukce dimenze</i>	5
Cézová Eliška	
<i>Ekonomicko-statistická optimalizace regulačního diagramu</i>	5
Cipra Tomáš	
<i>Některé kvantitativní aspekty penzí</i>	5
Dienstbier Jan	
<i>Covariate effects in extremes – remarks and theory</i>	6
Dohnal Gejza	
<i>Adaptivní regulační diagramy</i>	6
Donevska Sandra, Fišerová Eva, Hron Karel	
<i>Calibration between log-ratios of parts of compositional data using linear models</i>	6
Dvořák Jiří	
<i>Časoprostorové Coxovy bodové procesy s Lévyho bází</i>	7
Fabián Zdeněk	
<i>Resuscitace momentové metody</i>	7
Fačevicová Kamila	
<i>Použití logistické regrese pro diagnostiku výskytu rakoviny prostaty</i>	7
Fajfrová Lucie	
<i>Náhodná procházka na hierarchické grupě</i>	8
Fišerová Eva, Hron Karel	
<i>Ortogonalní regrese pro 3-složkové kompoziční data využitím lineárních modelů</i>	8
Friessl Michal	
<i>Testování normality ze zaokrouhlených dat</i>	9
Grendár Marian	
<i>Prevádzať p-hodnotu na Bayesov faktor?</i>	9
Hanel Martin, Buishand Adri	
<i>Regional block-maxima modelling of precipitation extremes in climate model simulations</i>	9
Hladík Milan, Černý Michal	
<i>Algoritmy, složitost a intervalová data</i>	10
Hlubinka Daniel, Kotík Lukáš	
<i>Zobecněná poloprostorová hloubka a její stejnoměrná silná konvergence</i>	11
Horáková Hana	
<i>Detekce změny ročního chodu průtokových řad</i>	11
Hron Karel	
<i>Předzpracování kompozičních dat</i>	11
Hrůzová Klára	
<i>Bilance a bilanční dendrogram kompozičních dat</i>	12
Hudecová Šárka	
<i>Testování změn v binárních autoregresních modelech</i>	12
Hušková Marie	
<i>Sekvenční testování stability ve funkcionálním modelu CAPM</i>	13
Chajdiak Jozef	
<i>Nezdanený objem hrubého domácího produktu, hrubý domácí produkt a daň z přidané hodnoty SR (rok 1996 až rok 2011)</i>	13
Chvosteková Martina	
<i>Simultánne testovanie strednej hodnoty a variancie normálneho rozdelenia</i>	13

Jarušková Daniela	
<i>Detekce změny kovariančního operátoru</i>	14
Jonathan Philip	
<i>Modelling covariate effects in extremes</i>	14
Jurczyk Tomáš	
<i>STATISTICA software</i>	14
Kalivodová Alžběta	
<i>Užití kompozičního biplotu při analýze medicínských dat</i>	15
Kasanický Ivan, Eben Kryštof	
<i>Identifikace netypického chování fotovoltaických elektráren</i>	15
Kaspaříková Nikola	
<i>Některé prostředky pro analýzu sekvencí</i>	16
Klaschka Jan	
<i>Podruhé o výpočtu Blakerova konfidenčního intervalu: Balíček BlakerCI a jiné resty</i>	16
Konár Ondřej, Brabec Marek, Kasanický Ivan, Malý Marek, Pelikán Emil	
<i>Optimalizace osazování odběrných míst inteligentními plynoměry</i>	17
Koróny Samuel, Hronec Štefan	
<i>Problematika riešenia efektívnosti verejných vysokých škôl na Slovensku</i>	17
Král Jan	
<i>Metodika komplexního návrhu regulačního diagramu</i>	17
Kulich Michal	
<i>Jednovýběrový vážený t-test pro pozorování s různými rozptyly</i>	18
Kynčlová Petra	
<i>Dirichletovo rozdělení vzhledem k Aitchisonově míře na simplexu</i>	18
Kyselý Jan, Píček Jan	
<i>Extreme value analysis in climatology</i>	18
Lachout Petr	
<i>Lineární regrese trochu jinak</i>	19
Laník Peter	
<i>Semi-parametrický přístup k odhadování koeficientů ARMA modelů časových řad</i>	19
Legát David	
<i>Statistická analýza obrazu a kontrola jakosti</i>	19
Lechnerová Radka, Lechner Tomáš	
<i>Analýza časových řad formální komunikace obcí</i>	20
Marek Patrice	
<i>Modelování a predikce výsledků hokejových zápasů</i>	20
Mesiar Radko	
<i>Kopule ako nástroj modelovania štruktúry stochastickej závislosti náhodných vektorov</i>	20
Michalíková Eva, Benáček Vladimír	
<i>The factors of growth of small family businesses. A robust estimation of the behavioral consistency in the panel data models</i>	20
Mlčúchová Dana	
<i>Aplikované statistické metody v analýzách onkologických dat zvířecích experimentů</i>	21
Nagy Stanislav	
<i>Konzistencia hĺbky funkcií</i>	21
Novák Petr	
<i>Regrese v modelech oprav</i>	21
Omelka Marek, Gijbels Irène, Veraverbeke Noël	
<i>Copule, parciální a podmíněné korelační koeficienty</i>	22
Pawlas Zbyněk	
<i>Odhad rozdělení latence odezvy neuronu</i>	22
Pešta Michal, Hudecová Šárka	
<i>Asymptotic consistency and inconsistency of the chain ladder</i>	23
Prášková Zuzana	
<i>Robustní monitorování stability modelu CAPM</i>	23
Prokešová Michaela	
<i>Statistická inference pro Coxovy bodové procesy s Lévyho bází</i>	23
Rosipal Roman	
<i>Multi-way data analysis for advanced physiological estimation of cognitive status</i>	24
Roth Martin	
<i>A regional peaks-over-threshold model in a non-stationary climate</i>	24

Sabolová Radka	
<i>Testy pre regresné kvantily založené na metóde sedlového bodu</i>	24
Sečkárová Vladimíra, Dedecius Kamil	
<i>Dynamic Bayesian estimation in diffusion networks</i>	25
Shokirov Bobosharif K.	
<i>A lower bound for the mixture parameter and its estimator</i>	25
Schindler Martin	
<i>How to choose threshold in a POT model?</i>	26
Schlesinger Pavel	
<i>Zpracování textu jako pomocník při detekci podvodného chování</i>	26
Starinská Katarína	
<i>Parameters estimates for change-point detection problem in AR time series</i>	26
Stibůrek David	
<i>Asymptotická ekvivalence statistik spojitého difúzního procesu pro náhodné časy</i>	27
Šedivá Blanka	
<i>Stabilita optimální volby portfolia Markowitzova modelu</i>	27
Tuček Pavel, Tučková Michaela, Harman Radoslav	
<i>Optimální návrh měření sigmoidálních funkcí</i>	27
Tučková Michaela, Kubáček Lubomír, Tuček Pavel	
<i>Design experimentu pro regresní modely s podmínkami typu I</i>	28
Vencálek Ondřej	
<i>Klasifikační metoda k nejbližších sousedů a hloubka dat</i>	28
Witkovský Viktor	
<i>Exaktné testy a konfidenčné oblasti pre parametre normálneho lineárneho modelu s dvomi variančnými komponentami</i>	28
Zikmundová Markéta, Staňková-Helisová Kateřina a Beneš Viktor	
<i>Užití částicového marginálního Metropolisova Hastingsova algoritmu ve stochastické geometrii</i>	29
Žambochová Marta	
<i>Modifikace algoritmu FEKM</i>	30

Jaromír Antoch**O segmentaci velmi dlouhých časových řad**

KPMS MFF UK v Praze

antoch@karlin.mff.cuni.cz

Cílem příspěvku bude diskuse o vybraných algoritmech umožňujících segmentaci velmi dlouhých časových řad, tj. řad obsahujících desetitisíce až milióny prvků. Ukážeme si přitom, že „úzkým hrdlem láhve“ je velikost dostupné operační paměti. V okamžiku, kdy je RAM vyčerpána a mezivýsledky se musí ukládat na pomalá média, veškeré doposud navržené algoritmy ztrácí velmi rychle na kvalitě a výkonosti.

Santiago Beguería Portugués**Covariate-dependent modelling of extreme events by nonstationary POT analysis**

CSIC, Zaragoza, Spain

santiago.begueria@csic.es

Non-stationary extreme value theory (NSEVT) is a relatively recent generalization of the extreme value theory allowing for varying (non-stationary) model parameters. In the NSEVT the pdf's of extreme values may vary as a function of covariates, establishing a functional relationship between them. The last years have seen an increasing number of studies applying NSEVT to climatic variables, helped by the development of new analysis resources. Most studies focused on identifying temporal trends in the occurrence of extreme events, i.e. having time as the covariate. However, nothing prevents applying NSEVT techniques to other covariates with an expected influence on the occurrence of extreme events such as meteorological indices. In this talk I will present ongoing research on the relationship between extreme precipitation and teleconnection indices in the Iberian Peninsula. After stating the main hypothesis and describing the datasets used I will present the non-stationary peaks-over-threshold (NSPOT) methodology used and the results of the analysis. I will then focus on the problems found during this research, and on alternative analysis options. The main purpose of the talk will be promoting the exchange of ideas and the discussion about unsolved or less clear aspects of NSPOT analysis.

Acknowledgement: The author and the research team KLIMATEXT benefited from project CZ.1.07/2.3.00/20.0086 (Strengthening international cooperation of the KLIMATEXT research team). The project CZ.1.07/2.3.00/20.0086 is co-financed by the European Social Fund and the state budget of the Czech Republic.

Jaromír Běláček**O vizualizaci statistických dat**

BioStat při Ústavu Biofyziky a Informatiky 1.LF UK Praha, VFN

jaromir.belacek@lf1.cuni.cz

Vizualizace dat je ve statistice dlouhodobě opomíjeným a zdánlivě zanedbatelným atributem prezentací výsledků formálních analýz. V reálných aplikacích však platí, že grafické zobrazení vysoce napomáhá přiblížení či pochopení podstaty či smyslu provedeného statistického testu – co není zobrazeno na grafu či na grafickém schématu zůstává neoborníkovi ve statistice buď špatně srozumitelné anebo často zcela nejasné. V podstatě nezáleží jen na tom, zda jde o aplikaci nejjednoduššího t -testu anebo o statistické úlohy složitější, které vyžadují skutečně preciznější vizualizované ošetření. Ve většině případů nám situaci neusnadňuje ani běžně dostupný statistický SW.

Cílem tohoto příspěvku je ukázat na několika systematicky vybraných příkladech z praxe: 1) které statistické testy můžeme z hlediska vizualizace považovat za „triviální (až zbytečné)“, 2) která „formálně triviální (až nesmyslná)“ zobrazení mohou být ve skutečnosti funkční, a 3) které „zdánlivě triviální“ úlohy mohou doopravdy vyžadovat pečlivou (předběžnou nebo posteriální) vizuální analýzu. Prostě a jednoduše: „Jde o to, aby si koncový uživatel odnesl od statistika opravdu to, co (by) si přál.“

Příkladem dobře ilustrujícím úlohy ad 1) i 2) může být standardní $X - Y$ graf s proloženou lineární přímkou, kde „efekt vizualizace“ závisí výhradně na volbě souřadnicových škál (a také na stupni standardizace proměnných X a Y). Typickým případem ad 3) může být zdánlivě jednoduchý model 2-Way ANOVA s jedním faktorem fixním a jedním faktorem opakovaní, kde lze smysluplně testovat fakticky 7 věcně relevantních hypotéz, nehledě na problém negaussovsky rozdělených dat.

Opravdu adekvátní grafická prezentace úlohy či výsledků analýzy může vyžadovat i mnoho hodin „optimalizační“ a také „tvůrčí“ práce. Na závěr příspěvku budou demonstrovány ukázky úspěšných (ale i ne zcela povedených) aplikací.

Sylvie Bělašková¹, Jan Janoušek² a Júlia Volaufová³

Optimal site for cardiac pacing in children

¹Tomas Bata University in Zlin, Czech Republic

²Children's Heart Centre, University Hospital Motol, Prague, Czech Republic

³Louisiana State University Health Sciences Center, New Orleans, USA

belaskova@fai.utb.cz

Background: We evaluated the effects of the site of ventricular pacing on left ventricular (LV) synchrony and function in children requiring permanent pacing.

Methods: 178 children (age to 18 years) from 21 Centres with atrioventricular block and structurally normal heart undergoing permanent pacing were cross-sectionally studied.

Conclusion: The site of ventricular pacing has a major impact on LV mechanical synchrony, efficiency and pump function in children that require life long pacing. LVA/LVLat pacing allows for optimal prevention of pacing-induced heart failure.

Viktor Beneš¹, Ondřej Šedivý¹ a Jakub Staněk²

Náhodné množiny s doprovodnými proměnnými a redukce dimenze

¹KPMS, MFF UK v Praze, ²KDM, MFF UK v Praze

benes@karlin.mff.cuni.cz

Redukce dimenze pro mnohorozměrná data byla rozvinuta Y. Guanem na prostorové bodové procesy s doprovodnými proměnnými. V příspěvku je toto dílo zobecněno v trojím smyslu. Rozšíření na náhodné kótované množiny je přímočaré. V metodě plátkované inverzní regrese je navrženo plátkování pomocí geometrických kót. Ve zjemněném modelu pro redukci dimenze je vyšetřován centrální podprostor druhého řádu. Simulační studie demonstruje navržené postupy.

Eliška Cézová

Ekonomicko-statistická optimalizace regulačního diagramu

ČVUT, FS, katedra matematiky, Karlovo nám. 13, 12000 Praha 2

eliska_c@email.cz

Statistická regulace je jednou z metod, jak provádět kontrolu procesu v reálném čase a reagovat na případné změny a poruchy. Tyto změny v procesu se snažíme korigovat a předcházet jim včasnou údržbou. Vhodně zvolená údržbová strategie může uspořit značné prostředky, které by musely být vynaloženy na neplánované opravy a seřízení výrobního procesu a s tím spojené ztráty z nekvalitní výroby v době, kdy se proces díky neočekávaným poruchám dostane do stavu mimo statistickou kontrolu. Na druhé straně je třeba počítat s náklady na údržbu už ve fázi návrhu statistické regulace procesu. Příspěvek se zabývá modely pro ekonomickou optimalizaci regulačního diagramu zahrnujícími různé typy údržby. Tyto modely jsou rozšířením dnes už klasického modelu, který představili v roce 1986 T.J. Lorenzen a L.C. Vance.

Tomáš Cipra

Některé kvantitativní aspekty penzí

Katedra pravděpodobnosti a matematické statistiky MFF UK, Praha 8

cipra@karlin.mff.cuni.cz

Penzijní aktiva se stávají významnou součástí osobních aktiv v moderní společnosti. V souvislosti s tím se objevuje celá řada důležitých kvantitativních aspektů, které je nutné matematicky analyzovat (a to zvláště v souvislosti s penzijní reformou v ČR). Příspěvek se zabývá některými aspekty tzv. rizika dlouhověkosti, jako jsou typy penzijních systémů, problém udržitelnosti penzí a podkladová úmrtnostní data včetně konkrétních aplikací pro Českou republiku.

Jan Dienstbier**Covariate effects in extremes – remarks and theory**

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

dienstbier.jan@gmail.com

We describe some theoretical aspects of extremal data modelling under the presence of covariate effects. We work with regression quantiles and analyze the simple situation of linear models $\mathbf{Y}_{n \times 1} = \mathbf{X}_{n \times p} \boldsymbol{\beta}_{p \times 1} + \mathbf{E}_{n \times 1}$ with the errors $E_i \sim F, i = 1, \dots, n$ fulfilling the domain of attraction condition. We show that the regression quantiles $\hat{\beta}_n(\alpha)$ itself share properties of the underlying extreme value distribution of the errors. Hence, we can develop tools on the basis of regression quantiles providing inferential methods dealing with extremal characteristics of linear models. These methods are based on Bahadur representations of the regression quantile process for $\alpha \in [0, 1]$. Our version of the Bahadur representation enables to approximate the tails of regression quantiles, i.e. regression quantiles $\hat{\beta}_n(\alpha_n)$ with α_n such that $\alpha_n n$ is an intermediate sequence in the terms of extreme value theory. The extremal inference is done on the base of smooth functionals of the tail quantile function, while the tail quantile function itself can be estimated using the Bahadur representation of regression quantiles. Various possible estimates of the empirical tail quantile functions are considered.

Acknowledgement: The author and the research team KLIMATEXT benefited from project CZ.1.07/2.3.00/20.0086 (Strengthening international cooperation of the KLIMATEXT research team). The project CZ.1.07/2.3.00/20.0086 is co-financed by the European Social Fund and the state budget of the Czech Republic.

Gejza Dohnal**Adaptivní regulační diagramy**

Fakulta strojní ČVUT v Praze

gejza.dohnal@csc-sro.cz

Při statistickém řízení procesů jsou aplikovány postupy a metody, které vyžadují poměrně přísné předpoklady pro jejich použití. V případě porušení těchto předpokladů jsou tyto metody neefektivní, vedou ke zvýšenému výskytu falešných signálů. Proto se hledají robustní verze regulačních diagramů, které jsou vůči porušení předpokladů méně citlivé. Robustní regulační diagramy však zpravidla prodlužují zpoždění při detekci vymezitelné příčiny. Tento negativní jev lze do jisté míry odstranit adaptivním přístupem.

Adaptivní sekvenční detekční schéma (SDS) reaguje na momentální stav procesu, odhadovaný na základě výsledků měření, prováděných v průběhu inspekce. Podle aktuální informace o procesu volíme parametry SDS pro další období (do další inspekce). V této souvislosti rozlišujeme mezi dvěma skupinami parametrů: parametry výběru a parametry diagramu. Adaptivním postupem lze do jisté míry napravit pomalou reakci například při použití robustního regulačního diagramu. Adaptivní SDS má za cíl především zkrátit zpoždění regulačního diagramu.

Další přínos adaptivního přístupu spočívá ve zpřesnění odhadů parametrů SDS. Množství dat, která máme k dispozici z fáze I SDS je zpravidla omezené a abychom získali potřebné množství hodnot, museli bychom vynaložit velké prostředky. Nedostatečný počet výběrů ve fázi I potom vede k velkým nejistotám při odhadu. Použitím naměřených dat ve 2. fázi SDS (při on-line regulaci) můžeme původní odhady postupně zpřesňovat adaptivním způsobem. Tento přístup opět vede k adaptivním regulačním diagramům.

Sandra Donevska, Eva Fišerová and Karel Hron**Calibration between log-ratios of parts of compositional data using linear models**

Palacký University Olomouc, Czech Republic

sdonevska@seznam.cz

Compositional data are multivariate observations carrying only relative information, popularly represented as proportions or percentages. Consequently, only ratios between parts of compositional data are informative [1, 4]. They are characterized by the simplex sample space with the Aitchison geometry that has Euclidean vector space structure. Thus, since compositional data have different nature from the standard multivariate observations that rely on the Euclidean geometry in real space, they need to be expressed in real space using proper log-ratio transformation before standard statistical analysis is applied.

In the contribution we will perform calibration between parts of compositional data. One possible way to solve this problem is to apply orthogonal regression to all log-ratios of pairs of compositional parts. We will focus on some properties and interpretation on matrices of predicted averages and residual variances as results for all the mentioned combinations of log-ratios. The corresponding statistical inference will be performed using a linear regression model with type-II constraints [2, 3].

References

- [1] Aitchison J. The statistical analysis of compositional data. Chapman and Hall, London, 1986.
- [2] Donevska S., Fišerová E. and Hron K. On the equivalence between orthogonal regression and linear model with type-II constraints. Acta Univ. Palacki. Olomuc., Fac. rer. nat. Math., 50, 19–27, 2011.
- [3] Fišerová E. and Hron K. Total least squares solution for compositional data using linear models. Journal of Appl. Statist. 37, 1137–1152, 2010.
- [4] Pawlowsky-Glahn V. and Buccianti A. Compositional data analysis: Theory and applications. Wiley, Chichester, 2011.

Jiří Dvořák

Časoprostorové Coxovy bodové procesy s Lévyho báží

KPMS MFF UK, Sokolovská 83, Praha 8; ÚTIA AV ČR, v.v.i., Pod Vodárenskou věží 4, Praha 8

`dvorak@karlin.mff.cuni.cz`, `dvorak@utia.cas.cz`

Coxův bodový proces, neboli dvojně stochastický proces, nabízí flexibilní rámec pro modelování shlukových bodových procesů. Bude představen model nestacionárního časoprostorového Coxova procesu využívající jádrové zhlazení Lévyho báze jako řídicí funkci intenzity a budou diskutovány možnosti odhadu parametrů modelu v případě časoprostorové separability použitého vyhlazovacího jádra. Přestože tento proces není separabilní, umožňuje splnění této podmínky založit odhady parametrů na marginálním časovém a prostorovém procesu, tedy projekci procesu do časové, resp. prostorové domény.

Zdeněk Fabián

Resuscitace momentové metody

Ústav informatiky AV ČR, v.v.i., Pod Vodárenskou věží 2, 18207 Praha 8

`zdenek@cs.cas.cz`

Buď $\mathcal{X} \subseteq \mathbb{R}$ otevřený interval a X_1, \dots, X_n iid výběr z rozdělení F , člena parametrické rodiny $\{F_\theta, \theta \in \Theta\}$ s nosičem \mathcal{X} a hustotami $f(x; \theta)$. Nejstaršími, zcela odlišnými metodami odhadu vektoru θ jsou momentová metoda a metoda maximální věrohodnosti. Maximálně věrohodné odhady jsou optimální, momenty jsou zato přímými charakteristikami datového souboru.

Buď S šikvná funkce. Pro $k \in \mathcal{N}$, k -tý moment náhodné veličiny $S(X)$ indexované vektorem θ je hodnota

$$ES^k(\theta) = \int_{\mathcal{X}} S^k(x; \theta) f(x; \theta) dx. \quad (1)$$

Momentová metoda odhaduje θ z rovnic

$$\frac{1}{n} \sum_{i=1}^n S^k(x_i; \theta) = ES^k(\theta), \quad k = 1, \dots, m, \quad (2)$$

představujících konečnou aproximaci (1). Ve vzorcích se odedávna používá funkce $S(x; \theta) = x$, což má za následek, že pro řadu rozdělení integrály (1) nekonvergují a metoda ztrácí smysl.

V přednášce zavedu šikvnější funkci, kterou nazývám skalární skór. Přesvědčení, že skalární skór je právě ona vhodná funkce, zakládám na tom, že pro libovolný \mathcal{X} je $S(x; \theta)$ pro určitou podmnožinu rozdělení na \mathcal{X} totožná se skórovou funkcí metody maximální věrohodnosti. V přednášce ukážu přednosti této volby: momenty existují, data lze popsat jejich výběrovými hodnotami, rovnice (2) buď dávají robustní odhady nebo je lze na takové poměrně snadno upravit.

Kamila Fačevicová

Použití logistické regrese pro diagnostiku výskytu rakoviny prostaty

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

`facza@seznam.cz`

V případě, kdy pracujeme s alternativně rozdělenou vysvětlovanou proměnnou, je vhodné použití modelu logistické regrese, viz [1]. Výstupem tohoto modelu však již není odhad hodnoty vysvětlované veličiny při daných

hodnotách proměnných vysvětlujících, nýbrž odhad pravděpodobnosti, že bude tato hodnota rovna jedné. Odhady parametrů tohoto modelu provádíme metodou maximální věrohodnosti. Protože jsou ale normální rovnice nelineární v parametrech, musíme je řešit některou z iteračních metod. Získané odhady parametrů potom vyjadřují logaritmus poměru šancí, že bude vysvětlovaná veličina rovna jedné v případě, kdy se příslušný regresor zvýší o jednotku a zbylé zůstanou neměnné.

Cílem příspěvku je za pomoci modelu logistické regrese a záznamů z vyšetření, která proběhla na Urologické klinice Fakultní nemocnice v Olomouci, odhalit faktory, které mají významný vliv na výsledek rebiopsie, tento vliv kvantifikovat a následně také porovnat s vlivem týchž faktorů na výsledek první biopsie, viz [2]. Výsledkem biopsie, resp. rebiopsie je v tomto případě myšleno, zda byl odhalen karcinom prostaty či nikoliv. Hlavní otázkou potom je, zda výsledek rebiopsie závisí na diagnóze, jež byla stanovena při biopsii první, viz [3].

Literatura

- [1] Agresti A. *Categorical Data Analysis*, second edition. John Wiley & Sons, Inc., 2002.
- [2] Fačevicová K. Použití logistické regrese pro diagnostiku výskytu rakoviny prostaty. Diplomová práce, Univerzita Palackého v Olomouci, 2012.
- [3] Grepl M., Študent V., Fürst T. a Fürstová J. Prostate cancer detection yield in repeated biopsies is independent of the diagnosis of earlier biopsies. *Biomedical papers* 4/2009, 297-305

Lucie Fajfrová

Náhodná procházka na hierarchické grupě

Ústav teorie informace a automatizace AV ČR, Pod Vodárenskou věží 4, 18207 Praha 8

fajfrova@utia.cas.cz

Jak je známo pro náhodnou procházku na \mathbb{Z}^d , to, zda bude střední doba návratu do počátku konečná či nekonečná, závisí na dimenzi d Euklidovského prostoru. Podobně je to i s problémem prvního setkání tří nezávislých náhodných procházek (míníme zde setkání libovolných dvou z nich) a (ne)konečnosti střední doby takového setkání. Právě tato úloha hraje roli při studiu kritické dimenze pro částicové systémy s větvením a anihilací, kterážto se ukazuje být někde mezi hodnotou jedna a dvě. Jenže dimenze Euklidovského prostoru je „jen“ celé číslo. Potřeba „neceločíselné dimenze“ nás vede ke studiu náhodné procházky na hierarchické grupě, která v jistém ohledu kopíruje chování náhodné procházky na \mathbb{Z}^d a jejíž parametr, analogický dimenzi, je číslo reálné (kladné). Naším motem je tedy nalezení „kritické dimenze“, kde se mění charakter chování tří náhodných procházek z „potkají“ na „nepotkají“. V tomto příspěvku ukážeme, jak asymptotické chování pravděpodobnosti, že se žádné dvě z procházek nepotkají do daného času, závisí na parametrech hierarchické grupy.

Příspěvek na základě společné práce s Janem Swartem a Noemi Kurt. Článek v přípravě.

Eva Fišerová a Karel Hron

Ortogonalní regrese pro 3-složkové kompoziční data využitím lineárních modelů

Katedra matematické analýzy a aplikací matematiky, Přírodovědecká fakulta Univerzity Palackého v Olomouci

eva.fiserova@upol.cz

Ortogonalní regrese je jedním z nástrojů pro statistické modelování dat, kdy se chyby vyskytují jak ve vysvětlované, tak i ve vysvětlující proměnné. V nejjednodušším případě se jedná o úlohu proložit n -tici dvourozměrných dat tak, aby součet čtverců vzdáleností jednotlivých bodů pozorování od odhadnuté přímky byl minimální. Ortogonalní regrese je tudíž invariantní vzhledem k rotaci souřadnic a lze ji aplikovat i při analýze vztahu mezi složkami kompozičních dat.

Cílem příspěvku je prezentovat iterační algoritmus pro odhad ortogonalní regresní přímky využitím lineárního modelu s podmínkami typu-II a ukázat možnosti provedení základních statistických inferencí. Teoretické výsledky budou aplikovány při analýze věkové struktury populace členských států OSN.

Literatura

- [1] Fišerová E. a Hron K. Total least squares solution for compositional data using linear models. *Journal of Applied Statistics* 37 (7), 1137–115, 2010.
- [2] Fišerová E. a Hron K. Statistical inference in orthogonal regression for three-part compositional data using a linear model with type-II constraints. *Communications in Statistics – Theory and Methods* 41 (13–14), 2367–2385, 2012.

Michal Friesl

Testování normality ze zaokrouhlených dat

Katedra matematiky FAV ZČU, Plzeň

friesl@kma.zcu.cz

Stojíme před úkolem provést test dobré shody s normálním rozdělením na základě pozorování seskupených do intervalů — pozorována jsou např. zaokrouhlená data, v našem případě celočíselná. Rozdělení může mít malý rozptyl (směrodatnou odchylku srovnatelnou s šíří intervalů) a rozsah výběru může být malý. V příspěvku budeme zvažovat použití variant Kolmogorovova-Smirnovova testu a chí-kvadrát testu dobré shody.

Marian Grendár

Prevádzka p-hodnotu na Bayesov faktor?

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

marian.grendar@savba.sk

Je známe, že p-hodnota nie je práve najšťastnejšie zvolenou mierou štatistickej evidencie. Niektorí Bayesiánci sa obetavo snažia p-hodnotu zachrániť prostredníctvom prevodu na Bayesov faktor. Pokúsime sa posúdiť zmysluplnosť takejto snahy.

Martin Hanel¹ and Adri Buishand²

Regional block-maxima modelling of precipitation extremes in climate model simulations

¹Technical University of Liberec, Czech Republic

²Royal Netherlands Meteorological Institute, De Bilt, Netherlands

Martin.Hanel@vuv.cz

The generalized extreme value (GEV) distribution has often been used to describe the distribution of daily maximum precipitation in observed and climate model data. A problem with extreme precipitation is that the likelihood of detecting a systematic change at a single station/grid box is generally small owing to the large year-to-year variability. The effect of this variability can be reduced using spatial pooling of the precipitation maxima over grid boxes.

The presented statistical model assumes that the precipitation maxima at each grid box follow GEV distribution. The model further allows the GEV location parameter to vary over the region, while the dispersion coefficient (the ratio of the GEV scale and location parameters) and the GEV shape parameter are assumed to be constant over the region. This corresponds with the index flood assumption in hydrology. It is further assumed that all three GEV parameters vary with time, such that the relative change in a quantile of the distribution is constant over the region. The uncertainty is assessed by a bootstrap resampling and the performance of the model is tested with the Anderson-Darling test.

The statistical model was applied to a number of regional climate model (RCM) simulations from the ENSEMBLES project aiming at the assessment of the performance of these RCM simulations in reproducing present climate and evaluation of the projected changes. The applications include the assessment of projected changes of 1-day summer and 5-day winter precipitation extremes over the Rhine basin, systematical evaluation of seasonal precipitation extremes of durations varying from 1 to 30 days for the Czech Republic and comparison of the RCM performance in simulation of 1-day and 1-hour precipitation extremes over the Netherlands. Finally, preliminary results from the analysis of observed subdaily precipitation extremes in the Czech Republic are mentioned. The performance of the RCMs in simulation of the 1-day precipitation extremes is in general reasonable, problems were identified for subdaily and multi-day precipitation extremes. Most of the RCM simulations project an increase in the annual and seasonal precipitation extremes.

Acknowledgement: The authors and the research team KLIMATEXT benefited from project CZ.1.07/2.3.00/20.0086 (Strengthening international cooperation of the KLIMATEXT research team). The project CZ.1.07/2.3.00/20.0086 is co-financed by the European Social Fund and the state budget of the Czech Republic.

Milan Hladík^{1,2}, Michal Černý²

Algoritmy, složitost a intervalová data

¹Univerzita Karlova v Praze, Matematicko-fyzikální fakulta, Katedra aplikované matematiky, Malostranské nám. 25, CZ 118 00 Praha, Česká republika

²Vysoká škola ekonomická v Praze, Fakulta informatiky a statistiky, nám. W. Churchilla 4, CZ130 67 Praha, Česká republika

milan.hladik@matfyz.cz, milan.hladik@vse.cz, cernym@vse.cz

Intervalová data se přirozeně vyskytují v řadě situací díky nejistotě, nepřesnosti měření nebo nedostatku informací. To je praktická motivace ke studiu zobecnění statistických pojmů a metod pro intervalová data. Přednáška se zaměří především na algoritmy a výpočetní složitost vybraných problémů motivovaných regresí s intervalovými daty. Složitostní klasifikace těchto problémů — typicky pomocí důkazu příslušnosti problému do třídy P či důkazu NP -těžkosti — ukazuje, pro které typy problémů existují efektivní algoritmy a pro které typy problémů není naděje na efektivní algoritmické řešení.

Když se podaří prokázat větu, že jistý problém je algoritmicky těžký (například NP -těžký), pak taková věta dává vysvětlení, proč je zajímavé zkoumat speciální případy (instance), ve kterých je problém algoritmicky efektivně zvládnutelný, a jak naopak vypadají ty instance, které činí problém v obecné formulaci obtížným.

Problém 1. Jako první ilustraci složitostního přístupu uvažme intervalovou matici \mathbf{X} a intervalový vektor \mathbf{y} . Množinu

$$\widehat{B}(\mathbf{X}, \mathbf{y}) := \{\widehat{\beta} \in \mathbb{R}^m : X^T X \widehat{\beta} = X^T \mathbf{y}, X \in \mathbf{X}, \mathbf{y} \in \mathbf{y}\}$$

lze chápat jako množinu všech možných odhadů parametrů regresního modelu

$$\mathbf{y} = X\beta + \varepsilon \quad (3)$$

metodou nejmenších čtverců, jestliže matice X probíhá intervalovou matici \mathbf{X} a vektor \mathbf{y} probíhá intervalový vektor \mathbf{y} . (Množina $\widehat{B}(\mathbf{X}, \mathbf{y})$ představuje jedno možné zobecnění pojmu „odhad modelu (3) metodou nejmenších čtverců“ pro případ, máme-li k dispozici jen intervalová data (\mathbf{X}, \mathbf{y}) .)

Ukážeme, že z algoritmického hlediska je množina $\widehat{B}(\mathbf{X}, \mathbf{y})$ „složitá“. Přesněji: ukážeme, že problém rozhodnout, zdali je množina $\widehat{B}(\mathbf{X}, \mathbf{y})$ neomezená, je NP -úplný. Předpokládáme-li $P \neq NP$, znamená to například, že neexistují efektivní algoritmy, které by dokázaly zkonstruovat intervalovou či elipsoidovou obálku množiny $\widehat{B}(\mathbf{X}, \mathbf{y})$.

Problém 2. Pro druhou ilustraci složitostního přístupu se zabýváme optimalizačním problémem

$$\min_{\beta \in \mathbb{R}^m} \|X\beta - \mathbf{y}\|. \quad (4)$$

Problém (4) je totiž typicky prostředkem k odhadu parametrů modelu (3). Zde $\|\cdot\|$ značí libovolnou vektorovou normu; zaměříme se ovšem především na L_p -normu s $p \in \{1, 2, \infty\}$.

Zabýváme se zobecněním problému (4) pro případ, kdy hodnoty matice X probíhají intervalovou matici \mathbf{X} a hodnoty vektoru \mathbf{y} probíhají intervalový vektor \mathbf{y} . Pak lze na (4) nahlížet jako na třídu optimalizačních problémů. Zajímá nás množina všech optimálních hodnot problému (4). Optimální hodnotu (4) nazveme *residuální hodnotou*. Supremum (infimum) z residuálních hodnot přes $X \in \mathbf{X}$ a $\mathbf{y} \in \mathbf{y}$ nazveme *největší (nejmenší) residuální hodnotou*.

Ukážeme, že

- pro *jakoukoli* normu $\|\cdot\|$ v (4) platí, že výpočet největší i nejmenší residuální hodnoty je NP -těžký;
- výpočet největší residuální hodnoty zůstává NP -těžký i v případě, kdy apriori víme, že regresní parametry jsou nezáporné;
- ovšem víme-li apriori, že regresní parametry jsou nezáporné, lze nejmenší residuální hodnotu spočítat efektivně pro p -normu s $p \in \{1, \infty\}$. Pro $p \in \{1, \infty\}$ ukážeme také několik odhadů na největší optimální hodnotu.

Problém 3. V problému 1 jsme zavedli množinu $\widehat{B}(\mathbf{X}, \mathbf{y})$ — je to vlastně množina všech optimálních řešení problému (4) při L_2 -normě. Zde se zabýváme její analogií s L_p -normou při volbě $p \in \{1, \infty\}$. Ukážeme, že i v tomto případě je její aproximace (např. intervalová či elipsoidová) NP -těžká. Protože v případě $p \in \{1, \infty\}$ je úloha (4) redukovatelná na lineární programování, ukážeme několik aplikací teorie lineárního programování s intervalovými daty. Zaměříme se na tzv. bazickou stabilitu (= existuje báze, která je optimální pro všechny realizace hodnot z intervalů \mathbf{X}, \mathbf{y}). Je-li model bazicky stabilní, pak množina optimálních řešení (a tím i optimálních hodnot) jde určit přesně a rychle. Ukážeme, že testování bazické stability je NP -těžké; nicméně existují pro ni efektivně testovatelné postačující podmínky.

Poděkování: Práce byla podpořena grantem GAČR P403/12/1947.

Daniel Hlubinka a Lukáš Kotík

Zobecněná poloprostorová hloubka a její stejnoměrná silná konvergence

KPMS MFF UK, Praha

hlubinka@karlin.mff.cuni.cz

Zobecněná poloprostorová hloubka byla vytvořena s cílem vylepšit některé nepříznivé vlastnosti poloprostorové hloubky. Zejména značný nesoulad mezi konturami hustoty a hloubky a to i v případě l_p symetrických rozdělání (s výjimkou sférických pro $p = 2$). Myšlenkou zobecnění je nahrazení klasického afinního poloprostoru zobecněným poloprostorem a zapojení váhových funkcí. Tím se dostáváme k metodě, která umí plynule propojit globální s lokálními, hloubku a hustotu.

V příspěvku se kromě základních vlastností zobecněné poloprostorové hloubky budeme zajímat o to, v čem je jsou silná a slabá místa této metody a hlavně se zaměříme na to, pro jaké váhové funkce je i zobecněná poloprostorová hloubka stejnoměrně silně konzistentní.

Hana Horáková

Detekce změny ročního chodu průtokových řad

Stavební fakulta ČVUT, Praha 6

horakovah@mat.fsv.cvut.cz

Časové řady průměrných denních průtoků se obvykle považují za stacionární. Přesto se zdá, že i v nich může docházet ke změnám, jestliže studujeme průměrné denní průtoky. V tomto příspěvku navrhneme statistické metody pro detekci změn v ročním cyklu. Metody jsou založeny na testování rovnosti středních hodnot dvou vektorů. Aplikovali jsme je na několik českých řek. Vzhledem k tomu, že délka studovaných řad byla 50-90 let, je síla testů bohužel poměrně malá. Použití testů stacionarity ročního cyklu bylo zamítnuto pouze u 3-5 řad.

Karel Hron

Předzpracování kompozičních dat

UPOL, PříF, KMAaAM, 17. listopadu 12, 771 46 Olomouc

hronk@seznam.cz

Kompoziční data představují mnohorozměrná pozorování vyjadřující kvantitativní popisy relativních příspěvků částí na celku, s četnými aplikacemi v geologii, analytické chemii, biologii, medicíně nebo ekonomice [6]. Ve starší literatuře se o kompozičních datech často hovoří jako o datech s konkrétním konstantním součtem složek (rovným 1 v případě proporcí a 100 u procentuálních podílů), ten ovšem představuje pouze vhodnou reprezentaci kompozic; fixace na určitý předepsaný součet složek vede naopak k paradoxům a obecně k nerozumným výsledkům [1]. Specifické vlastnosti kompozičních dat indukují tzv. Aitchisonovu geometrii na simplexu, výběrovém prostoru kompozic, se strukturou euklidovského vektorového prostoru [2, 3]. Hlavním metodickým nástrojem statistické analýzy kompozičních dat je jejich vyjádření v souřadnicích vzhledem k vhodné zvolené ortonormální bázi na simplexu prostřednictvím izometrických logratio transformací [2, 3]. Vzhledem k výskytu logaritmu podílů kompozičních složek v uvedené třídě transformací je potřeba před samotnou statistickou analýzou kompozičního datového souboru vedle eliminace dalších artefaktů, jako například výskytu chybějících hodnot, také ošetřit přítomnost nulových hodnot v datech. Řešení obou zmíněných problémů je na rozdíl od standardních mnoho-rozměrných pozorování ztíženo skutečností, že veškerá informace u kompozičních dat je obsažena v podílech mezi jejich složkami, což by měla každá odpovídající imputační metoda přirozeně respektovat. Cílem příspěvku je popsat metody nahrazení chybějících hodnot a tzv. zaokrouhlovacích nul (vzniklých jako hodnot pod mezí detekce měřícího přístroje) pomocí iteračního algoritmu, využívajícího regresní modelování a specifickou interpretaci ortonormálních souřadnic [4, 5]. Teoretické výsledky budou podpořeny výstupy simulační studie, případně též demonstrací na reálných datech.

Literatura

- [1] Aitchison J. The statistical analysis of compositional data. Chapman & Hall, London, 1986.
- [2] Egozcue J. J., Pawłowsky-Glahn V., Mateu-Figueras G. a Barceló-Vidal C. Isometric logratio transformations for compositional data analysis. *Mathematical Geology* 35(3), 279–300, 2003.
- [3] Hron K. Elementy statistické analýzy kompozičních dat. *Informační Bulletin ČStS* 21(3), 41–48, 2010.
- [4] Hron K., Templ, M. a Filzmoser P. Imputation of missing values for compositional data using classical and robust methods. *Computational Statistics & Data Analysis* 54(12), 3095–3107, 2010.

- [5] Martín-Fernández J. A., Hron K., Templ M., Filzmoser P. a Palarea-Albaladejo J. Model-based replacement of rounded zeros in compositional data: classical and robust approaches. *Computational Statistics & Data Analysis* 56(9), 2688–2704, 2012.
- [6] Pawlowsky-Glahn V., Buccianti, A. eds. *Compositional data analysis: theory and applications*. Wiley, Chichester, 2011.

Klára Hrůzová

Bilance a bilanční dendrogram kompozičních dat

Univerzita Palackého v Olomouci, Přírodovědecká fakulta, 17. listopadu 12 771 46 Olomouc

klara.h.87@centrum.cz

Kompoziční data jsou mnohorozměrná data, nesoucí pouze relativní informaci, zajímají nás tedy podíly mezi jednotlivými složkami kompozičního vektoru spíše než (absolutní) hodnoty složek jako takových [1]. Výběrovým prostorem reprezentací kompozičních dat s konstantním součtem (procenta, proporce) je simplex s odpovídající, Aitchisonovou, geometrií. Vzhledem k poměrně složité interpretaci výsledků analýzy kompozičních dat na simplexu a nemožnosti použít standardní statistické metody (opírajících se o euklidovskou geometrii) v tomto výběrovém prostoru byly zavedeny logratio transformace kompozičních dat do reálného prostoru. Jednou z nich je izometrická logratio (ilr) transformace generující souřadnice vzhledem k ortonormální bázi na simplexu, díky které zobrazíme D -složkové kompozice ze simplexu do reálného prostoru dimenze $D - 1$ s euklidovskou geometrií [3].

Pro konstrukci ortonormální báze na simplexu můžeme využít tzv. postupné binární dělení, které volíme zejména z důvodu dobré interpretace výsledných souřadnic (tzv. bilancí) v ilr transformaci [2]. Název bilance odpovídá vyjádření souřadnic, jedná se vlastně o poměr mezi skupinami složek, které vznikly v postupném binárním dělení. Můžeme se tedy následně zajímat jak o vztahy mezi těmito skupinami (meziskupinová analýza), nebo o vztahy v dané skupině (vnitroskupinová analýza). Pro průzkumovou statistickou analýzu bilancí, přiřazených kompozičnímu datovému souboru, využíváme nástroj zvaný bilanční dendrogram [4]. Tento budeme aplikovat na reálných datech z ekonomiky, přičemž budeme podrobně analyzovat také otázku vhodné volby postupného binárního dělení.

Literatura

- [1] Aitchison J. *The statistical analysis of compositional data*. Chapman & Hall, London, 1986.
- [2] Egozcue J. J. a Pawlowsky-Glahn V. Groups of parts and their balances in compositional data analysis. *Mathematical Geology* 37, 795–828, 2005.
- [3] Egozcue J. J., Pawlowsky-Glahn V., Mateu-Figueras G. a Barceló-Vidal C. Isometric logratio transformations for compositional data analysis. *Mathematical Geology* 35, 279–300.
- [4] Thió-Henestrosa S., Egozcue J. J., Pawlowsky-Glahn V., Kovács L. Ó. a Kovács G. P. Balance-dendrogram. A new routine of CoDaPack. *Computers & Geosciences* 34, 1682–1696, 2008.

Šárka Hudecová

Testování změn v binárních autoregresních modelech

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

hudecova@karlin.mff.cuni.cz

Časové řady binárních náhodných veličin hrají důležitou roli v celé řadě praktických aplikací. Typicky se s nimi setkáváme při sledování výskytu určité události v čase, jako např. denní výskyt srážek, měsíční nebo čtvrtletní indikátory recese aj. V literatuře bylo navrženo několik různých přístupů k modelování takových časových řad. Jedním z nejrozšířenějších jsou tzv. binární autoregresní (BAR) modely, které kombinují principy klasických autoregresních modelů časových řad a zobecněných lineárních modelů, viz [1]. V tomto příspěvku se budeme zabývat testováním přítomnosti změny v parametrech BAR modelů. Navržený postup vychází z metodologie a výsledků [2].

Literatura

- [1] Kedem B. a Fokianos K. *Regression models for time series analysis*. Wiley, New York, 2002.
- [2] Antoch J., Gregoire G., a Jarůšková D. Detection of structural changes in generalized linear models. *Stat. and Probab. Lett.* 69, 315–332, 2004.

Marie Hušková

Sekvenční testování stability ve funkčním modelu CAPM

MFF UK Praha

huskova@karlin.mff.cuni.cz

Príspevek se týká sekvenčních procedur pro detekci nestability v parametrech modelu CAPM (capital asset pricing model), jestliže máme k dispozici tzv. vysokofrekvenční data. Z hlediska statistického se jedná o problém detekce nestability ve speciálním mnohorozměrném regresním modelu.

Bude pojednáno o konstrukci testových statistik i jejich limitních vlastnostech data. Teoretické výsledky budou aplikovány na reálná data.

Literatura

- [1] Aue A., Hoermann S., Horváth L., Hušková M. a Steinebach J. Sequential testing for the stability of high frequency portfolio betas. *Econometric Theory* 28, 804–837, 2012.

Jozef Chajdiak

Nezdanený objem hrubého domáceho produktu, hrubý domáci produkt a daň z pridanej hodnoty SR (rok 1996 až rok 2011)

Slovenská štatistická a demografická spoločnosť

chajdiak@statis.biz

Dane existujú snád' aj dlhšie ako ľudstvo a slúžia na krytie kolektívneho uspokojovanie potrieb zdaňovaného priestoru. Častý je prístup deficitného hospodárenia. Autor oproti tomu považuje nerovnosť **príjmy** \geq **výdavky** za axiomatickú ekonomickú požiadavku.

Ukážeme si, že objem nezdaneného HDP, čo je niečo vyše 30 mld. EUR, predstavuje priestor na zvýšenie príjmov, lebo predstavuje pri zdanení vyše 2 mld. EUR DPH navyše. Stav v štátnom rozpočte SR, že v roku 2012 treba znížiť schodok o 300 mil. EUR a v roku 2013 aspoň o 1,5 mld. EUR má tak svoj priestor.

Literatura

[1] www.statistics.sk

[2] www.finance.gov.sk

[3] Pohľady na ekonomiku Slovenska 2001 (až 2012). Bratislava, SSDS 2001 (až 2012) – elektronická verzia na www.ssdssk

Pod'akovanie: Vypracované v rámci riešenia úlohy VEGA č. 1/1164/12 „Možnosti uplatnenia informačných a komunikačných technológií na zvyšovanie efektívnosti medzinárodnej spolupráce malých a stredných podnikov SR v oblasti inovácií.“

Martina Chvosteková

Simultánne testovanie strednej hodnoty a variancie normálneho rozdelenia

Ústav merania SAV, Bratislava

chvosta@gmail.com

V príspevku sa zaoberáme simultánnym testovaním strednej hodnoty μ a variancie σ^2 normálneho rozdelenia. Pre uvažovanú hypotézu $H_0 : (\mu, \sigma^2) = (\mu_0, \sigma_0^2)$ neexistuje rovnomerne nejsilnejší test. V literatúre možno nájsť viacero približných ale i presných testov použitých aj na konštrukciu oblasti spoľahlivosti pre oba neznáme parametre zároveň. Porovnali sme sily presných testov (Mood, 1950; Choudhari–Kundu–Misra, 2001) a približných testov navrhnutých v Arnold–Shavelle (1998). Spočítali sme aj obsahy oblastí spoľahlivosti skonštruovaných jednotlivými testami, pričom bližné testy sme modifikovali.

Pod'akovanie: Práca vznikla vďaka podpore grantov APVV-0096-10, VEGA 2/0019/10 a VEGA 2/0038/12.

Literatúra

- [1] Arnold B. C. a Shavelle R. M. Joint confidence sets for the mean and variance of a normal distribution. *The American Statistician* 52, 133–140, 1998.
- [2] Choudhari P., Kundu D. a Misra N. Likelihood ratio test for simultaneous testing of the mean and variance of a normal distribution. *Journal of Statistical Computation and Simulation* 71, 313–333, 2001.
- [3] Mood A. M.: *Introduction to the theory of statistics*. New York, McGraw-Hill, 1950.

Daniela Jarušková

Detekce změny kovariančního operátoru

Stavební fakulta ČVUT, Praha 6

jarus@mat.fsv.cvut.cz

Příspěvek navazuje na přednášku na Robustu 2010 a na článek Panaretos, Kraus and Maddocks: Second order comparison of Gaussian random functions and the geometry of DNA minicircles, publikovaném v JASA (2010), v kterém se autoři zabývají dvouvýběrovým testem o shodnosti dvou kovariančních operátorů. V článku se však explicitně nedefinuje, proti jaké alternativě je test uvažován. V mém příspěvku ukazuji, že přirozenou alternativou by mohla být hypotéza o rozdílnosti K -dimenzionální aproximací (ve smyslu hlavních komponent) obou operátorů. V tom případě je však nutno volit jinou testovou statistiku, o které lze ukázat, že má za platnosti nulové hypotézy standardní normální rozdělení a za platnosti alternativy je konzistentní. Bohužel pro konečný počet dat má navrhaná testová statistika velké vychýlení. Pro získání přibližných kritických hodnot navrhuje použití permutační princip.

Philip Jonathan

Modelling covariate effects in extremes

Shell Projects and Technology & Lancaster University, UK

philip.jonathan@shell.com

Extreme value analysis can help us understand unusual events in our physical environment, particularly given current concerns about climate change, providing a mathematically sound and statistically efficient basis for modelling. For example, reliable design and assessment of flood and coastal defences and marine structures requires estimation of both marginal and dependence characteristics of extreme environments. Incorporation of covariate effects is necessary for good modelling. For example, by expressing the parameters of extreme value distributions as smooth functions of storm direction, we can model the directional variation of extreme ocean storms. Similarly we can estimate seasonal, temporal and spatial variation. Characterisation of dependence structure is also critical for good modelling of joint occurrences of rare events. For example, in a spatial context, rare events are often spatially clustered. The most extreme environmental loads on a marine structure may correspond to joint occurrence of large waves, winds and currents. This talk will illustrate some current methodologies for covariate and dependence modelling in extreme value analysis in application to extreme ocean environments.

Acknowledgement: The authors and the research team KLIMATEXT benefited from project CZ.1.07/2.3.00/20.0086 (Strengthening international cooperation of the KLIMATEXT research team). The project CZ.1.07/2.3.00/20.0086 is co-financed by the European Social Fund and the state budget of the Czech Republic.

Tomáš Jurczyk

STATISTICA software

StatSoft CR, Ringhofferova 115/1, CZ 155 21 Praha 5 - Zličín

tomas.jurczyk@statsoft.cz

Software *STATISTICA* je komplexním systémem pro zpracování a analýzu dat používaný v mnoha odvětvích. Analytické nástroje tohoto softwaru pokrývají širokou škálu oblastí statistiky od popisné statistiky přes metody pro řízení kvality až po sofistikované data-miningové metody. Cílem prezentace je představit účastníkům tento statistický software, který na MFF není příliš známý, ačkoli je v praxi běžně využíván jak v Čechách, tak i ve světě. Na praktickém příkladě si ukážeme, jak se v softwaru pracuje, a řekneme si, jaké jsou jeho výhody a nevýhody.

Alžběta Kalivodová

Užití kompozičního biplotu při analýze medicínských dat

Laboratoř dědičných metabolických poruch Lékařská fakulta, Univerzita Palackého v Olomouci
UPOL, PřF, KMAaAM, 17. listopadu 12, 771 46 Olomouc

Kalivodovaa@gmail.com

Biplot je v současnosti hojně užívaný grafický nástroj mnohorozměrné statistické analýzy [1]. Jedná se o rovinný graf, jehož konstrukce vychází ze singulárního rozkladu datové matice, respektive příslušných skórá a zátěží prvních dvou hlavních komponent a slouží k zachycení mnohorozměrné datové struktury (obvykle znázorněna pomocí bodů) a vztahů mezi proměnnými (šipky). Často se proto aplikuje také při statistické analýze speciálních typů dat, tzv. kompozičních dat, nesoucích pouze relativní informaci (speciálně procenta, proporce) [2]. Při jejich statistické analýze se užívá rodina tzv. logratio transformací, která umožňuje zpracování kompozičních dat standardními statistickými metodami. Pro konstrukci kompozičního biplotu je výhodné aplikovat tzv. centrovanou logratio transformaci, která indukuje jeho výhodnou interpretaci [2]. Tato bude porovnána s interpretací biplotu pro standardní mnohorozměrná pozorování (tj. nesoucích absolutní informaci). Oba uvedené druhy biplotů budou v tomto příspěvku aplikovány na reálných datech. Neprve budou pro ilustraci představena data obsahující relativní zastoupení onemocnění jako příčin úmrtí ve vybraných zemích Evropy. Dále budou použita data reprezentující krevní vzorky pacientů s různými dědičnými metabolickými poruchami týkající se metabolismu aminokyselin a acylovaných karnitinů. Tato data budou srovnána s kontrolními vzorky. Vzhledem ke skutečnosti, že pro chemometrická data je biplot konstruován na základě velkého počtu proměnných (chemických látek), budou v grafickém výstupu vynechány šipky reprezentující tyto proměnné (zátěže) a zůstanou zobrazena pouze pozorování (pacienti a kontroly) v podobě bodů (skóry) s cílem vyšetřit datovou strukturu, zejména výskyt shluků.

Literatura

- [1] Aitchison J. The statistical analysis of compositional data. Chapman & Hall, London, (1986).
- [2] Aitchison J. a Greenacre M. Biplots of compositional data. Journal of the Royal Statistical Society, 51(4), 375–392, 2002.
- [3] Gabriel, K.R. The biplot graphic display of matrices with application to principal component analysis. Biometrika, 58(3), 453–467, 1971.

Ivan Kasanický a Kryštof Eben

Identifikace netypického chování fotovoltaických elektráren

Ústav informatiky AV ČR, v.v.i., Pod Vodárenskou věží 2, 18207 Praha 8

kasanicky@cs.cas.cz

V průběhu několika posledních let proběhl v České republice takzvaný solární boom a v současné době je v ČR zapojeno více než 13 000 fotovoltaických elektráren (FVE) s celkovým instalovaným výkonem téměř 2 MWp. FVE patří mezi neregulované zdroje a proto musí být elektřina z jejich produkce plně vyčerpána dříve než se začne využívat energie z ostatních (neobnovitelných) zdrojů. Proto je potřeba v každé chvíli vědět co nejpřesněji, kolik energie fotovoltaické zdroje právě dodávají do sítě. Z různých důvodů však nejsou zdaleka všechny FVE osazeny dálkově odečítaným průběhovým měřením, a tak se celková výroba všech fotovoltaických zdrojů v ČR počítá pomocí extrapolace výroby měřených zdrojů.

Z tohoto důvodu je třeba identifikovat případné netypické chování některé z měřených FVE, jako je například částečná odstávka elektrárny, a zohlednit tuto informaci při výpočtu celkové výroby. Tento úkol je však ztížený faktem, že výroba FVE je určena intenzitou slunečního záření, oblačností a dalšími meteorologickými veličinami. Zejména sluneční záření a oblačnost přitom mají velkou volatilitu.

V příspěvku budou představeny metody detekce netypického chování fotovoltaických elektráren. Tyto metody budou založeny nejen na porovnání výrob jednotlivých fotovoltaických elektráren mezi sebou, ale také na zkoumání funkcionálního vztahu mezi sluneční radiací naměřenou satelity nebo pozemními stanicemi a elektrickou energií vyrobenou jednotlivými fotovoltaickými zdroji. Tento postup by měl umožnit sledovat a srovnávat proces výroby mezi jednotlivými farmami, ale také odhadovat změnu chování určité farmy.

Nikola Kaspářková**Některé prostředky pro analýzu sekvencí**

Vysoká škola ekonomická v Praze

school@tulipany.cz

V souvislosti s rozvojem technologií pro získávání dat jsou stále dostupnější také datové soubory, které mají podobu kategoriálních sekvencí. Vedle sekvencí sledovaných v biologii (třeba aminokyseliny v bílkovinách) mohou být zajímavá i data o chování subjektů ve společenskovedních nebo ekonomických aplikacích. Příkladem analýzy sekvencních dat jsou dnes již běžné analýzy chování uživatelů na webu prováděné nad souborem záznamů o webových stránkách odeslaných z daného serveru. Při řešení úloh v praxi může být cílem najít vhodně shrnující popis chování zkoumaných subjektů a vybrat ze souboru typické sekvence. Pozornost se věnuje metodám určení reprezentativních sekvencí a v souvislosti s tím postupům pro zjištění podobnosti dvou sekvencí, případně volbě substituční matice. Jednotlivé přístupy se liší tím, co výsledek vyjadřuje, a výpočetní náročností, která je u rozsáhlejších souborů podstatná.

Jan Klaschka**Podruhé o výpočtu Blakerova konfidenčního intervalu:****Balíček BlakerCI a jiné resty**

Ústav informatiky AV ČR, Pod Vodárenskou věží 2, CZ–18207 Praha 8

klaschka@cs.cas.cz

O výpočtu Blakerova konfidenčního intervalu jsem mluvil už na ROBUSTu 2010 [6]. Prezentoval jsem tehdy základní představu o algoritmu, který by se měl vypořádat s nedostatky původního Blakerova algoritmu ze základní práce [1] (včetně opravy [2]).

Připomeňme si, že

- Blakerův konfidenční interval pro parametr p binomického rozdělení 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 $\{p; \beta(p) > \alpha\}$, kde β je tzv. funkce přijatelnosti (acceptability function),
- komplikace při numerickém výpočtu těchto mezí plynou z toho, že funkce β je spojitá jen po částech a ve spojitých úsecích nemusí být monotónní.

Myšlenka algoritmu, který je založen na analýze vlastností funkce β , byla v době konání ROBUSTu 2010 sotva několik týdnů stará. Není tedy divu, že se řada věcí kolem algoritmu udála až později; o některých z nich budu na ROBUSTu 2012 referovat.

Na prvním místě půjde o balíček `BlakerCI` [7] v R. Na pořad by mělo přijít také srovnání s „konkurenčním“ algoritmem M. P. Faye [4, 5]. (Fayův algoritmus je daleko sofistikovanější než Blakerův, ale má i své slabiny.) Dále bych se mohl zmínit (bude-li ovšem kdy) např. o radostech počítání v aritmetice s omezenou přesností, když funkce β je stejně jako závislost Blakerových konfidenčních mezí na parametru α nespojitá.

Literatura

- [1] Blaker H. Confidence curves and improved exact confidence intervals for discrete distributions. *Canadian J. of Statistics* 28, 783–798, 2000.
- [2] Blaker H. Corrigenda: Confidence curves and improved exact confidence intervals for discrete distributions. *Canadian J. of Statistics* 29, 681, 2001.
- [3] Clopper C. J. a Pearson E. S. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* 26, 404–413, 1934.
- [4] Fay M. P. Confidence intervals that match Fisher's exact and Blaker's exact tests. *Biostatistics* 11, 373–374, 2010.
- [5] <http://cran.r-project.org/web/packages/exactci>
- [6] Klaschka J. O výpočtu Blakerova konfidenčního intervalu. ROBUST 2010. Sborník abstraktů, 29. ČStS Praha 2010.
- [7] <http://cran.r-project.org/web/packages/BlakerCI>

Ondřej Konár, Marek Brabec, Ivan Kasanický, Marek Malý a Emil Pelikán

Optimalizace osazování odběrných míst inteligentními plynoměry

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

konar@cs.cas.cz

Celosvětovým trendem v oblasti měření spotřeby plynu je postupné osazování odběrných míst tzv. inteligentními měřidly. Tyto přístroje jednak měří ve vysokém časovém rozlišení a jednak umožňují on-line přenos naměřených dat ke zpracování v informačním systému distributora nebo obchodníka s plynem. Ačkoli cena těchto přístrojů postupně klesá, vzhledem k velmi vysokému počtu odběrných míst (např. v ČR jich existuje přes milion) je nutné osazování provádět postupně v průběhu několika let.

Ideální je rozmisťovat měřidla tak, aby byla naměřená data využita s maximální efektivitou. Vzhledem k tomu, že u zákazníků bez průběhového měření je denní spotřeba odhadována modelem, je například vhodné přednostně osazovat zákazníky s vyšší variabilitou odběru, u nichž se dá předpokládat vyšší chyba odhadu.

V příspěvku bude představena metodika výběru vhodných odběrných míst k osazení inteligentním měřením. Metodika je založena na statistické analýze fakturačních údajů odběrných míst ze zákaznického kmene distribuční společnosti RWE GasNet, s.r.o. Hlavními kritérii jsou přitom variabilita odběru individuálních zákazníků v čase a zastoupení „problematických“ zákazníků v jednotlivých obcích. Prezentovaná metodika bude v uvedené distribuční společnosti v následujícím roce provozně testována.

Samuel Koróny a Štefan Hronec

Problematika riešenia efektívnosti verejných vysokých škôl na Slovensku

Univerzita Mateja Bela, Banská Bystrica

samuel.korony@umb.sk

Autori v príspevku uvádzajú vybrané výsledky analýz urobených v rámci projektu VEGA 1/0969/11 „Matematicko-ekonomické metódy hodnotenia efektívnosti verejných vysokých škôl na Slovensku“. Po nutnom úvode do problematiky merania a hodnotenia efektívnosti produkčných jednotiek sú v prvej časti opísané dva základné analytické postupy - parametrický štatistický (na základe produkčných funkcií) a neparametrický DEA (pomocou lineárneho programovania). V druhej časti sú uvedené konkrétne výsledky aj v podobe poradia verejných vysokých škôl pre niektoré vstupy a výstupy úrovně vzdelávania a vedy a techniky.

Jan Král

Metodika komplexního návrhu regulačního diagramu

ISQ PRAHA s.r.o.

kral.jan@seznam.cz

Tato práce řeší problematiku komplexního přístupu k návrhu regulačního diagramu a řeší etapu návrhu teoreticky správného efektivního a v reálné praxi implementovatelného typu regulačního diagramu prostřednictvím nově vyvinuté metodiky včetně navržených metodických schémat.

Předkládaná práce byla navržena a řešena jako reakce na konkrétní aktuální požadavky strojírenských podniků, které požadují průkazné, jednoznačné a rychlé řešení vznikajících problémů za pomoci SW podpor a jen nezbytně nutných finančních nákladů. Dále byla požadována garance dostatečné vědecko-teoretické úrovně a metodické podpory při praktickém zavádění nově navržené metodiky.

Vědecko-teoretickou úroveň lze doložit aktivní účastí a autorskými výstupy z řešení projektu zadaného grantem 1M06047 – Centrum pro jakost a spolehlivost výroby vypsáním Ministerstvem školství, mládeže a tělovýchovy v letech 2006–2011 a praktickou implementací v strojírenských podnicích, zejména v GCE Chotěboř s.r.o.

Práce se z důvodu rozsahu předmětné problematiky zabývá újeji problematikou analýzy systému měření a ekonomicko-statistickými aspekty v návrzích regulačního diagramu. Pro tyto oblasti byly zpracovány výpočtové šablony pro MS Excel na podporu analýzy systému měření MSA a vedení regulačních diagramů s rozšířeními mezemi, které jsou přílohou této práce. Navržené hypotézy byly v průběhu řešení ověřeny a předložená práce potvrzuje jejich oprávněnost a význam pro teorii i praxi, prioritně v podmínkách strojírenského podniku.

Implementace navrženého postupu pro návrh a zavedení SPC představuje rovněž jednu z nezbytných podmínek pro dosažení a udržení konkurenceschopnosti produkce podniku, zejména v současných náročných podmínkách globalizace.

Michal Kulich

Jednovýběrový vážený t-test pro pozorování s různými rozptyly

KPMS MFF UK, Sokolovská 83, 186 75 Praha 8

kulich@karlin.mff.cuni.cz

Klasický t-test předpokládá nezávislá stejně rozdělená pozorování. V našem příspěvku ukážeme, jak zobecnit klasický t-test na pozorování se stejnou střední hodnotou a potenciálně různými rozptyly. Použijeme vážený průměr místo klasického a odvodíme momentový odhad rozptylu čitatele testové statistiky, který nevyžaduje znalost rozptylů jednotlivých pozorování ani schopnost je konsistentně odhadnout. Dokážeme asymptotickou normalitu odvozené testové statistiky a navrhneme aproximaci asymptotického rozdělení t-rozdělením, které lépe funguje při malých rozsazích výběru. Navržený test je vhodný pro porovnání středních hodnot veličin, které vznikly jako odhady z různých počtů pozorování učiněných uvnitř experimentálních jednotek.

Petra Kynčlová

Dirichletovo rozdělení vzhledem k Aitchisonově míře na simplexu

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

kynclova.petra@gmail.com

Kompoziční data jsou speciálním typem mnohorozměrných dat, která kvantitativně popisují části nějakého celku, tedy nesou pouze relativní informace [1]. Většina standardních statistických metod předpokládá, že zkoumaná data pocházejí z reálného prostoru s euklidovskou geometrií. Geometrická struktura simplexu, výběrového prostoru kompozičních dat, je přitom odlišná a je charakterizována tzv. Aitchisonovou geometrií. Proto se snažíme kompozice z Aitchisonovy geometrie zobrazit do standardní euklidovské geometrie, využíváme k tomu vyjádření kompozičních dat v souřadnicích vzhledem k vhodné zvolené ortonormální bázi na simplexu prostřednictvím izometrických logratio transformací [2]. V reálném prostoru s euklidovskou geometrií jsou hustoty rozdělení pravděpodobnosti vyjádřeny vzhledem k Lebesgueově pravděpodobnostní míře. Z tohoto důvodu se pro simplexový výběrový prostor zavádí alternativní, relativní míra. Tato míra se označuje jako Aitchisonova míra a je zavedena pomocí transformace Lebesgueovy míry z prostoru ortonormálních souřadnic na simplex [3, 4].

Jako vhodný nástroj pro parametrické modelování kompozičních dat se tradičně uvádí Dirichletovo rozdělení, jelikož předpokládá simplex jako výběrový prostor. Jeho hustota je ovšem typicky vyjádřena vzhledem k Lebesgueově míře. Cílem příspěvku je popsat vlastnosti a číselné charakteristiky Dirichletova rozdělení na simplexu vzhledem k Aitchisonově míře, resp. vzhledem k Lebesgueově míře v prostoru ortonormálních souřadnic, a porovnat je s vlastnostmi tzv. normálního rozdělení na simplexu. Zejména budou studovány důsledky volby parametrů na tvar Dirichletova rozdělení a případné možnosti využití tohoto rozdělení ve statistické analýze kompozičních dat.

Literatura

- [1] Aitchison J. The statistical analysis of compositional data. Chapman & Hall, London, 1986.
- [2] Egozcue J. J., Pawlowsky-Glahn V., Mateu-Figueras G., Barceló-Vidal C. Isometric logratio transformations for compositional data analysis. *Mathematical Geology*, 35(3), 279–300, 2003.
- [3] Mateu-Figueras G., Pawlowsky-Glahn V., Egozcue J. J. The principle of working on coordinates. In: V. Pawlowsky-Glahn, A. Buccianti (eds.) *Compositional data analysis: theory and applications*, pp. 31–42, Wiley, Chichester, 2011.
- [4] Monti G. S., Mateu-Figueras G., Pawlowsky-Glahn V., Egozcue, J. J. The shifted-scaled Dirichlet distribution in the simplex. In: J. J. Egozcue, R. Tolosana-Delgado, M. I. Ortego (eds.) *Compositional Data Analysis Workshop – CoDaWork’11, Proceedings, International Center for Numerical Methods in Engineering (CIMNE), Barcelona, (2011)*.

Jan Kyselý^{1,2} and Jan Píček¹

Extreme value analysis in climatology

¹Technical University of Liberec, Czech Republic

²Institute of Atmospheric Physics AS CR, Prague

kysel@ufa.cas.cz, jan.picek@tul.cz

We present a brief introduction to applications of extreme value analysis in climatology. Importance of the topic stems from the fact that extremes are associated with major effects on environment and society, and climate

change impacts will probably be manifested mainly through changes in extremes. Typical applications of extreme value analysis in climatology and related fields involve estimation of return levels of (observed or simulated) extremes, estimation of design values of meteorological variables (e.g. surface air temperature, precipitation amounts of various durations from minutes to days) needed in engineering practice (including design of structures such as nuclear power plants, dams, urban drainage systems etc.), estimation of changes/trends in extremes, and evaluation of changes in extremes as projected by climate models. In addition to “routine” methods (such as the block maxima method with the Generalized Extreme Value distribution), which do not require additional settings and are sometimes used as “black-box” tools to provide basic answers to basic questions, “advanced” methods of the extreme value analysis have undergone rapid development recently, with some of the advances motivated by climate research needs. These include extreme value models that incorporate covariates/non-stationarity and spatial and multivariate models for extremes.

Presentations within the KLIMATEXT session will introduce some of the recent advances in methods of the extreme value analysis in climatology and issues that are dealt with, including those to which answers are still sought.

Acknowledgement: The authors and the research team KLIMATEXT benefited from project CZ.1.07/2.3.00/20.0086 (Strengthening international cooperation of the KLIMATEXT research team). The project CZ.1.07/2.3.00/20.0086 is co-financed by the European Social Fund and the state budget of the Czech Republic.

Petr Lachout

Lineární regrese trochu jinak

KPMS MFF UK

lachout@karlin.mff.cuni.cz

Představíme model lineární regrese trochu jinak. Vyjdeme z posloupnosti pozorování a přiřadíme k nim teoretický, asymptotický model na základě četností. Procedura odhadování bude založena na měření velikosti reziduí a hledání epsilon-optimálního řešení. Tento přístup zahrnuje, jak model s deterministickými regresory, tak model s náhodnými regresory.

Peter Laník

Semi-parametrický přístup k odhadování koeficientů ARMA modelů časových řad

KM FPV, UMB, Tajovského 40, SK-97401 Banská Bystrica

Peter.Lanik@umb.sk

Pri odhadovaní koeficientov ARMA modelov stacionárnych časových radov sa klasické odhadovacie metódy opierajú o predpoklad znalosti pravdepodobnostného rozdelenia šumu, väčšinou gaussovského. V príspevku je prezentovaný prístup, ktorý sa neobmedzuje na jedno konkrétne rozdelenie, ale pripúšťa, že dáta môžu pochádzať aj z iného rozdelenia. Cieľom prístupu je robustnosť voči nesprávnej špecifikácii modelu z hľadiska rozdelenia šumu a zároveň voči aditívnym outlierom.

David Legát

Statistická analýza obrazu a kontrola jakosti

KPMS MFF UK v Praze

david.legat@gmail.com

Jednou z významných oblastí manipulace s nestrukturovanými daty je zpracování signálů jako je zvuk a obraz, pro které existuje velké množství postupů. Příspěvek se zabývá statistickým přístupem ke zpracování obrazu, při kterém je obraz interpretován jako reprezentant náhodného pole. Budou zmíněny dva problémy: odstranění šumu z obrazu, které napomáhá lepší interpretaci obrazu, a klasifikace obrazu, při které se snažíme identifikovat a rozpoznávat zobrazované objekty. Část příspěvku zaměřená na odstranění šumu pojednává především o využití simulačních metod MCMC, především pak o tzv. perfektních simulacích. V části příspěvku pojednávajícím o klasifikaci obrazu budou zmíněny různé modifikace metody klasifikačních stromů. Na závěr bude uveden příklad zpracování obrazu v kontrole jakosti, kde bude cílem identifikace vad tkaných textilií.

Radka Lechnerová¹ a Tomáš Lechner²

Analýza časových řad formální komunikace obcí

¹SVŠES, s.r.o., Lindnerova 575/1, 180 00 Praha 8-Libeň, ²VŠE v Praze, Národohospodářská fakulta, katedra práva, nám. W. Churchilla 4, 130 67 Praha 3

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

Komunikace je základem všech sociálních interakcí a je jednou z nejkomplesnějších aktivit vůbec. Obecný rozvoj informační společnosti má významný vliv na změny způsobů komunikace, přičemž tyto změny jsou reflektovány také veřejnou správou v rámci procesu implementace informačních a komunikačních technologií označovaných jako e-Government. Právní základ fungování veřejné správy dává také povinnost evidence formální komunikace, kterou můžeme díky této poměrně přesné evidenci zpětně analyzovat. Získané výsledky lze využít v rámci zpětné vazby v procesu reformy veřejné správy, která je dnes velmi významným evropským tématem. V příspěvku se zabýváme analýzou časových řad formální komunikace vybraných obcí ČR. Získané výsledky by měly umožnit lépe porozumět postupným změnám ve způsobech komunikace orgánů veřejné správy a izolovat od sebe různé vlivy, jejichž konkrétní působení je podstatné pro přípravu implementace dalších nástrojů e-Governmentu.

Patrice Marek

Modelování a predikce výsledků hokejových zápasů

Západočeská univerzita v Plzni

patrke@kma.zcu.cz

Modelování a predikce výsledků hokejových zápasů není tak prozkoumávanou oblastí jako modelování a predikce výsledků fotbalových zápasů. Cílem této práce je prozkoumat možnost použití modelů běžně používaných pro fotbalové zápasy, kde je předpokládáno, že výsledky mohou být modelovány pomocí dvourozměrného Poissonova rozdělení (případně vhodně modifikovaného). Konečný model rovněž zohledňuje čas, kdy byly jednotlivé zápasy odehrány.

Pro odhady parametrů modelů jsou dostupná data za více než 10 sezon v NHL (National Hockey League) a Extralize českého hokeje. Testování je pak provedeno na poslední dostupné sezoně (2010/2011). Celkové výsledky jsou porovnány s reálnými výsledky a s výsledky předpokládanými sázkovými kanceláři.

Radko Mesiar

Kopule ako nástroj modelovania štruktúry stochastickej závislosti náhodných vektorov

STU Bratislava

mesiar@math.sk

Po historickom úvode a uvedení filozofie pojmu kopule v 2- i n -rozmernom prípade rozoberieme Sklarovu vetu, vrátane podrobného netriviálneho príkladu. Uvedieme viaceré príklady kopúl, napr. Archimedovské, EV (Extreme Value), USC (Univariate Conditioning Stable) či Archimax kopule, vrátane rôznych konštrukčných metód. Pre dvojrozmerné kopule stručne uvedieme niektoré koeficienty asociácie, napr. Spearmanovo rho či Kendallovo tau.

Eva Michalíková^{1,2} a Vladimír Benáček^{2,3}

The factors of growth of small family businesses. a robust estimation of the behavioral consistency in the panel data models

¹Brno University of Technology, Faculty of Business and Management

²Charles University, Institute of Economic Studies, Prague

³SOÚ Academy of Sciences of the Czech Republic, Prague

michalikova@volny.cz, benacek@fsv.cuni.cz

The paper quantifies the role of factors associated with the growth (or decline) of micro and small businesses in European economies. The growth is related to employment and value added in enterprises as well as to ten institutional variables. We test the data for consistency of behavioural patterns in various countries and gradually remove outlying observations, quite a unique approach in the panel data analysis, that can lead to erroneous conclusions when using the classical estimators. In the first part of this paper we outline a highly robust method of estimation based on fixed effects and least trimmed squares (LTS). In its second part we apply this method on the panel data of 28 countries in 2002–2008 testing for the hypothesis that micro and small businesses in Europe use

different strategies for their growth. We run a series of econometric tests where we regress employment and total net production in micro and small businesses on three economic factors: gross capital returns, labour cost gaps in small relative to large enterprises and the GDP per capita. In addition, we also test the role of 10 institutional factors in the growth of family businesses.

Dana Mlčúchová

Aplikované statistické metody v analýzách onkologických dat zvířecích experimentů

LEM - ÚMTM, LF UP Olomouc

dana.mlcuchova@gmail.com

V příspěvku budou prezentovány statistické metody, které se rutinně používají při vyhodnocování protinádorového účinku léčiv. Základní výzkum účinku léčiv na zvířatech je v medicíně prvotním krokem k získání znalostí o chování léčivých látek v těle živých organismů. Při prokázání účinku léčiva následuje preklinický výzkum na tzv. buněčných liniích a posledním krokem je klinický výzkum. Při testování změn po podání léčiva ve vztahu k neléčené skupině se z důvodu častého porušení normality dat využívají zejména neparametické metody. Běžně používanou metodou je také analýza přežívání, protože významným ukazatelem účinku léčiva je přežívání zvířat v průběhu terapie.

Stanislav Nagy

Konzistencia hlĺbky funkcí

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

nagy@karlin.mff.cuni.cz

Hĺbka dat je nástrojom analýzy mnohorozmerných (a nekonečnerozmerných) dat so širokými možnosťami využitia v neparametrickej štatistike. Základnou vlastnosťou, ktorú rozumná hlĺbka musí splĺnať je rovnomerná konzistencia výberovej verzie. Bez nej nie je možné zabezpečiť esenciálne vlastnosti ako konzistenciu mnohorozmernej analógie mediánu (bodu s najväčšou hlĺbkou) alebo kontúr hlĺbky.

V príspevku sa zameriavame na funkcionálne dáta a niekoľko rôznych prístupov k určovaniu hlĺbky funkcí používaných v literatúre. Ukazujeme, že silný výsledok autorov López-Pintado a Romo [2, Theorem 4] o rovnomernej konzistencii výberovej pásovej hlĺbky *neplatí* (a to ani za silnejších predpokladov) a tvrdenie ilustrujeme na jednoduchých protipríkladoch. Ukazujeme, prečo dôkaz zlyháva a navrhujeme dva rôzne prístupy k úprave pásovej hlĺbky tak, aby rovnomernú konzistenciu už bolo možné zaručiť. Dokazujeme najsilnejšiu, uniformnú konzistenciu modifikovaných verzí hlĺbky. V prípade hlĺbok integrálneho typu tým zároveň rozširujeme známe výsledky Fraimana a Munizovej [1, Theorem 3.1].

Literatúra

- [1] Fraiman R. a Muniz G. Trimmed means for functional data. *Test* 10(2), 419–440, 2001.
- [2] López-Pintado S. a Romo J. On the concept of depth for functional data. *J. Amer. Statist. Assoc.* 104(486), 718–734, 2009.

Petr Novák

Regrese v modelech oprav

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

novakp@karlin.mff.cuni.cz

Při provozu systému který podléhá opotřebení je naší snahou odhadnout rozdělení doby do selhání pro optimalizaci plánování údržby. Pomocí vhodných modelů chceme popsat závislost tohoto rozdělení na případných regresorech. Běžně používané modely analýzy přežití, jako je Coxův model nebo model zrychleného času, je potřeba přizpůsobit systému s opravami, lze např. modelovat závislost na předchozích opravách. V příspěvku takové modely popisujeme a předvádíme využití na datech z praxe.

Marek Omelka¹, Irène Gijbels² a Noël Veraverbeke³

Copule, parciální a podmíněné korelační koeficienty

¹Charles University in Prague, Czech Republic; ²Katholieke Universiteit Leuven, Belgium; ³Hasselt University, Belgium

omelka@karlin.mff.cuni.cz

Předpokládejme, že pozorujeme nezávislé trojice $(\mathcal{X}_1, Y_{11}, Y_{21}), \dots, (\mathcal{X}_n, Y_{1n}, Y_{2n})$ z nějakého rozdělení (\mathcal{X}, Y_1, Y_2) . V následujícím nás bude zajímat vztah (závislost) dvojice reálných náhodných veličin (Y_1, Y_2) v situaci, kdy je zapotřebí vzít v úvahu také pomocnou náhodnou veličinu (kovariátu) \mathcal{X} , která může být jednodimenzionální, vícedimenzionální nebo dokonce funkcionální.

Za tímto účelem byla vyvinuta řada metod, která se liší svými předpoklady o vlivu \mathcal{X} na odezvu (Y_1, Y_2) (parciální korelační koeficienty, parametrické modely, ...). Velmi obecný přístup k této problematice představil [1]. Označme si sdružené podmíněné rozdělení (Y_1, Y_2) při daném $\mathcal{X} = \chi$ jako

$$H_\chi(y_1, y_2) = P(Y_1 \leq y_1, Y_2 \leq y_2 | \mathcal{X} = \chi)$$

a odpovídající marginální rozdělení pomocí

$$F_{1\chi}(y_1) = P(Y_1 \leq y_1 | \mathcal{X} = \chi), \quad F_{2\chi}(y_2) = P(Y_2 \leq y_2 | \mathcal{X} = \chi).$$

Pokud jsou podmíněná rozdělení $F_{1\chi}$ a $F_{2\chi}$ spojitá, potom dle Sklarova věty (Sklar's theorem) existuje jednoznačně daná funkce C_χ , která z podmíněných marginálních rozdělení vytváří sdružené podmíněné rozdělení:

$$H_\chi(y_1, y_2) = C_\chi(F_{1\chi}(y_1), F_{2\chi}(y_2)), \quad (y_1, y_2) \in \mathbb{R}^2.$$

Funkce C_χ plně popisuje vztah mezi (Y_1, Y_2) při daném $\mathcal{X} = \chi$ a nazývá se podmíněná kopule (conditional copula). Za povšimnutí stojí, že koncept podmíněné kopule je velmi obecný a dovoluje, že se závislost mezi Y_1 a Y_2 může měnit s hodnotou kovariáty $\mathcal{X} = \chi$.

Dále, jelikož se pořadové korelační koeficienty dají vyjádřit jako funkcionály kopule [2], lze pomocí podmíněné kopule definovat podmíněné „pořadové korelační koeficienty“, jako například:

$$\begin{aligned} \text{podmíněné Kendallovo tau:} \quad & \tau(\chi) = 4 \iint C_\chi(u_1, u_2) dC_\chi(u_1, u_2) - 1, \\ \text{podmíněné Spearmanovo rho:} \quad & \rho(\chi) = 12 \iint C_\chi(u_1, u_2) du_1 du_2 - 3. \end{aligned}$$

V příspěvku představíme neparametrický odhad podmíněné kopule C_χ a podmíněných korelačních koeficientů a jejich využití při zkoumání vztahu dvojice náhodných veličin.

Literatura

- [1] Patton, J. A. Modeling asymmetric exchange rate dependence. *International Economic Review* 47(2), 527–556, 2006.
 [2] Nelsen, R. B. An introduction to copulas. Springer, New York. Second edition, 2006.

Zbýněk Pawlas

Odhad rozdělení latence odezvy neuronu

KPMS MFF UK

pawlas@karlin.mff.cuni.cz

Uvažujme dva nezávislé náhodné výběry X_1, \dots, X_n a Y_1, \dots, Y_n , ze kterých pozorujeme vždy jen menší hodnotu $Z_i = \min(X_i, Y_i)$, $i = 1, \dots, n$. Naším cílem je odhadnout rozdělení náhodných veličin X_i . Jedná se vlastně o problém náhodného cenzorování s tím, že indikátory cenzorování nejsou známy. Místo toho máme k dispozici náhodný výběr $\tilde{Y}_1, \dots, \tilde{Y}_m$ ze stejného rozdělení jako mají náhodné veličiny Y_i .

Motivace pro studium tohoto problému pochází z neurofyzologie. Informace v nervovém systému je přenášena posloupností akčních potenciálů (tzv. spiků) generovaných neurony. Označme spiky generované za přítomnosti spontánní aktivity jako spontánní. Zajímá nás reakce neuronu na vnější stimulus (např. zvukový podnět). Ta se projeví jedním nebo více evokovanými spiky. Latence odezvy neuronu je doba od začátku stimulu po výskyt prvního evokovaného spiku. V praxi však naměříme dobu do výskytu prvního spiku po stimulu, ten může být buď spontánní, nebo evokovaný. Přitom nejsme schopni rozlišit, zda první pozorovaný spik byl způsoben spontánní aktivitou nebo reakcí na podnět. Znamená to, že pozorujeme minimum (Z_i) doby po první evokovaný spik (X_i)

a doby po první spontánní spike (Y_i). Odhad rozdělení latence odezvy založíme na pozorováních Z_i a časech \tilde{Y}_i , které mají stejné rozdělení jako Y_i a získáme je z měření během spontánní aktivity neuronu.

Michal Pešta and Šárka Hudecová

Asymptotic consistency and inconsistency of the chain ladder

Charles University in Prague, Faculty of Mathematics and Physics, Department of Probability and Mathematical Statistics, Czech Republic

pesta@karlin.mff.cuni.cz, hudecova@karlin.mff.cuni.cz

The distribution-free chain ladder reserving method belongs to the most frequently used approaches in the general insurance. It is well known, see [1], that the estimators \hat{f}_j of the development factors are unbiased and mutually uncorrelated under some mild conditions on the mean structure and under the assumption of independence of the claims in different accident years. We deal with some asymptotic properties of \hat{f}_j . Necessary and sufficient conditions for asymptotic consistency of the estimators of true development factors f_j are provided [2]. A rate of convergence for the consistency is derived. Possible violation of these conditions and its consequences are discussed, and some practical recommendations are given. Numerical simulations and a real data example are provided as well.

Literatura

- [1] Mack, T. Distribution-free calculation of the standard error of chain ladder reserve estimates. *ASTIN Bulletin* 23(2), 1993, 213–225.
- [2] Pešta, M. and Hudecová, Š. Asymptotic consistency and inconsistency of the chain ladder. *Insurance: Mathematics and Economics*, to appear, 2012.

Zuzana Prášková

Robustní monitorování stability modelu CAPM

MFF UK Praha

praskova@karlin.mff.cuni.cz

Oceňovací model kapitálových aktiv CAPM je známý model teorie portfolia, který vyjadřuje vztah rizika jednotlivých aktiv k tržnímu riziku. V tomto příspěvku je CAPM uvažován jako mnohorozměrný regresní model s časově proměnlivými parametry. Pro testování stability parametrů modelu v čase je navržen sekvenční postup, který je na rozdíl od obvykle používané metody nejmenších čtverců založen na M-odhadech a částečných součtech vážených M-reziduí. Bylo odvozeno asymptotické chování testové statistiky za nulové i alternativní hypotézy pro případ, že chyby i regresory jsou slabě závislé náhodné veličiny a vektory. V příspěvku budou prezentovány základní teoretické výsledky spolu s výsledky simulační studie a aplikací na reálná data.

Literatura

- [1] Chochola O., Hušková M., Prášková Z., Steinebach J. Robust monitoring of CAPM portfolio betas, zasláno, 2012.

Michaela Prokešová

Statistická inference pro Coxovy bodové procesy s Lévyho bází

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

prokesov@karlin.mff.cuni.cz

Coxovy bodové procesy s Lévyho bází představují flexibilní třídu modelů pro modelování shlukových (čas)prostorových bodových procesů. Takzvaná shot-noise reprezentace umožňuje vyjádření těchto procesů jako zobecněných shlukových procesů. Díky této reprezentaci je také jednoduché zahrnout do modelu nestacionaritu.

Obdobně jako i pro jiné modely založené na Lévyho bázích je pro tyto bodové procesy typické, že jejich věrohodnostní funkce je příliš složitá na to, aby se dala použít pro inferenci, zato momentové míry mají jednoduchý tvar. Odhady parametrů jsou proto založeny na momentových charakteristikách a metodě minimálního kontrastu. Příspěvek se bude zabývat odhady parametrů pro stacionární i nestacionární Coxovy bodové procesy s Lévyho bází a rovněž otázkami ověření odhadnutého modelu pomocí reziduí pro prostorové bodové procesy.

Roman Rosipal**Multi-way data analysis for advanced physiological estimation of cognitive status**

Institute of Measurement Science, Slovak Academy of Sciences, Bratislava

rosipal@savba.sk

We applied and developed new multi-way data analysis algorithms for physiological advanced estimation of cognitive status, which significantly improved the estimation of cognitive workload and shed new light on the estimation of mental fatigue. More specifically, we used atomic decomposition of parallel factor analysis (PARAFAC) and multi-way partial least squares (N-PLS) to identify unique sources of brain electrical activity as measured by the EEG recorded in human participants as they performed tasks that induced different mental states, including engagement, mental workload, and mental fatigue. We tested two types of atomic decomposition, each of which identifies unique EEG sources simultaneously in three dimensions: 1) atoms with dimensions of power spectral density, space (electrode position), and time (time on task or task conditions), or 2) atoms with dimensions of magnitude squared coherence, spatial relationships (electrode pairs), and time.

For tasks that induced mental workload, we found atoms that combine sources in the theta and alpha EEG frequency bands consistently in individual participants at different times of day and on different days. The temporal variations of the atoms clearly reflected the levels of mental workload induced by varying task conditions. For a task that induced mental fatigue, we found atoms that tracked the development of mental fatigue in individual participants over time, while reflecting underlying changes in power or coherence of primarily theta-band EEG. Our results show that atomic decomposition is a valuable new approach to the identification and measurement of EEG sources for monitoring cognitive status. By comparing these results with results of prior analyses using the same data sets, we observed that atomic decomposition can supplement or overcome existing approaches based on conventional two-dimensional space-time or frequency-time decomposition of EEG.

Acknowledgement: This work was carried out with Leonard J. Trejo (Pacific Development and Technology, LLC) and funded by ARO-W911NF-11-C-0081, ARO-W911NF-08-C-012, APVV-0096-10 and VEGA 2/0019/10.

Martin Roth**A regional peaks-over-threshold model in a non-stationary climate**

Royal Netherlands Meteorological Institute, De Bilt, Netherlands

martin.roth@knmi.nl

Regional frequency analysis is often used to reduce the uncertainty in the estimation of distribution parameters and quantiles. In this paper a regional peaks-over-threshold model is introduced that can be used to analyze precipitation extremes in a changing climate. We use a temporally varying threshold, which is determined by quantile regression for each site separately. The marginal distributions of the excesses are described by generalized Pareto distributions (GPD). The parameters of these distributions may vary over time and their spatial variation is modeled by the index flood (IF) approach. We consider different models for the temporal dependence of the GPD parameters. Parameter estimation is based on the framework of composite likelihood. Composite likelihood ratio tests that account for spatial dependence are used to test the significance of temporal trends in the model parameters and to test the IF assumption. We apply the method to gridded, observed daily precipitation data from the Netherlands for the winter season. A general increase of the threshold is observed, especially along the west coast and northern parts of the country. Moreover, there is no indication that the ratio between the GPD scale parameter and the threshold has changed over time, which implies that the scale parameter increases by the same percentage as the threshold. These positive trends lead to an increase of rare extremes of on average 22% over the country during the observed period.

Acknowledgement: The authors and the research team KLIMATEXT benefited from project CZ.1.07/2.3.00/20.0086 (Strengthening international cooperation of the KLIMATEXT research team). The project CZ.1.07/2.3.00/20.0086 is co-financed by the European Social Fund and the state budget of the Czech Republic.

Radka Sabolová**Testy pre regresné kvantily založené na metóde sedlového bodu**

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

sabolova@karlim.mff.cuni.cz

Metóda sedlového bodu poskytuje presné aproximácie hustoty odhadov aj pre malé rozsahy výberov. V príspevku sa budeme zaoberať testovou štatistikou, ktorá je založená na výraze v exponente v aproximácii hustoty pre

M-odhady. Vďaka tvaru funkcie ψ pre regresné kvantily získame explicitný vzorec pre testovú štatistiku. Odvožené testy sú asymptoticky ekvivalentné s klasickými testami založenými na vierohodnosti, no ich relatívna chyba je iba rádu $O(n^{-1})$. Správanie navrhnutých parametrických i neparametrických testov pri rôznych rozsahoch výberov a rozdeleniach bude ilustrované v simulačnej štúdii.

Vladimíra Sečkárová¹ a Kamil Dedecius²

Dynamic Bayesian estimation in diffusion networks

¹MFF UK, KPMS, Sokolovská 83, CZ–186 75 Praha 8

² ÚTIA AV ČR, v.v.i., Pod Vodárenskou věží 4, Praha 8

seckarov@karlin.mff.cuni.cz

The common use of ad-hoc networks and their growing complexity brings the question of reliable distributed estimation of variables describing the environment of interest. One of the commonly used methods is a centralized approach, where the network nodes communicate their data with a single specialized point. However, this method suffers from high communication overheads and represents a potentially dangerous concept with a single point of failure needing special treatment. Our aim is to contribute to another quite recent method called diffusion estimation. Here, the operating environment is decentralized; the network nodes communicate just within a close neighbourhood. For the modelling and estimation the Bayesian framework is adopted. Unlike in the traditional approaches, in each case a particular model is taken into account. This leads to a very scalable and universal method, applicable to a wide class of different models. As an example, we show application of the method to a selected member of the exponential family.

Bobosharif K. Shokirov

A lower bound for the mixture parameter and its estimator

CHMI, Na Šabatce 17, 143 06 Praha 4 and KPMS MFF UK, Sokolovská 83, 186 75 Praha 8

bobosari@karlin.mff.cuni.cz

With a sample X_1, \dots, X_n of size n drawn from a distribution function (d.f.) $H(x)$

$$H(x) = \theta F(x) + (1 - \theta)G(x), \quad x \in [0, 1], \quad (\theta \in (0, 1)), \quad (1)$$

where $F(x)$ and $G(x)$ are defined and continuous on the interval $[0, 1]$ d.f.'s, we study the problem of estimating θ , the mixture parameter. Based on approach and results of [1], the following result is obtained.

Theorem 1 *Let X_1, \dots, X_n be sample of size n drawn from d.f. $H(x)$, Y_1, \dots, Y_n be its transformed sample that corresponds to d.f. $1 - (1 - H(x))/(1 - F(x))$ and $1 \leq k \leq n$. Assume the following conditions hold:*

$$G(x) > F(x), \quad \forall x \in [0, 1], \quad (2)$$

$$S_G \subset [0, 1 - \delta], \quad \text{for some } \delta > 0, \quad (3)$$

and

$$\frac{F'(x)}{1 - F(x)} \leq \frac{G'(x)}{1 - G(x)}. \quad (4)$$

Let $\varphi(x)$ be a strictly decreasing on the interval $[0, 1]$ function, such that $\varphi(0) = -\varphi'(0) = 1$ and satisfies the relation

$$\frac{d^2}{dx^2} \left[\varphi^{-1} \left(\frac{1 - H(x)}{1 - F(x)} \right) \right] \geq 0. \quad (5)$$

Then for the mixture parameter in the model (1) the inequality

$$\theta \geq 1 - \frac{H(X) - F(X)}{F(X)(1 - \varphi(YR_H(y_0)))} \quad (6)$$

holds and its lower bound can be estimated as

$$\theta_n^* = \max \left\{ 1 - \frac{k}{n[1 - \varphi(YR_n(y_0))]}, 0 \right\}, \quad (7)$$

where Y is defined by

$$\max \{Y_1, \dots, Y_k\} \leq Y \leq \min \{Y_{k+1}, \dots, Y_n\}, \quad k \leq n, \quad (8)$$

$y_0 \in (0, Y)$, $x_0 \in (0, 1)$ is such that $\overline{H(y_0)} \cdot \overline{F(x_0)} = \overline{H(x_0)}$,

$$R_n(y_0) = \frac{1}{y_0} \varphi^{-1} \left(\frac{1 - H_n(x_0)}{1 - F(x_0)} \right)$$

and $H_n(x)$ is the empirical d.f., constructed by the sample X_1, \dots, X_n . θ_n^* , defined by (7), is an estimator for the mixture parameter θ in the model (1).

Literatura

- [1] Klebanov, L. B., Yakovlev, A. A. A new approach to testing for sufficient follow-up in cure rate analysis. *Journal of Statistical Planning and Inference* 137, 3557–3569, 2007.

Martin Schindler

How to choose threshold in a POT model?

FP TUL, KAP, Studentská 2, CZ – 461 17 Liberec
martin.schindler@tul.cz

The peaks-over-threshold (POT) method with a nonstationary threshold for estimating high quantiles (return levels) is investigated. It was shown that using (95%) regression quantile as the time-dependent threshold instead of a constant threshold can be beneficial. It is assumed that a linear trend is present in the data and so a linear regression quantile as the threshold is used. The aim is to find the threshold (regression quantile) which would be optimal with respect to the reliability of the estimates of high quantiles by means of Monte Carlo simulations. Based on this criterion stationary and regression quantile thresholds are compared. It is described how the choice of the optimal threshold depends on the sample size, estimated quantile or the estimate itself.

Acknowledgement: The authors and the research team KLIMATEX benefited from project CZ.1.07/2.3.00/20.0086 (Strengthening international cooperation of the KLIMATEX research team). The project CZ.1.07/2.3.00/20.0086 is co-financed by the European Social Fund and the state budget of the Czech Republic.

Pavel Schlesinger

Zpracování textu jako pomocník při detekci podvodného chování

SAS Institute CR, s.r.o., Na Pankráci 17-19, 140 21 Praha 4
pavel.schlesinger@cze.sas.com

V pojišťovnictví a bankovníctví je dnes již zavedenou částí práce vyšetřovatelů na detekci pojistných či interních podvodů zaměstnanců. Úspěšná detekce přináší nezanedbatelné snížení finančních nákladů. V posledních letech ke zlepšeným výsledkům vyšetřování dopomáhá automatické zpracování dat, novinkou přitom je analýza nestrukturovaných informací ukrytých v textu. Příspěvek přiblíží základy automatického procesu detekce podvodu a využití textu. Navíc budou prezentovány nástroje SASu využívané vyšetřovateli při vyhledávání relevantního textu a vizualizaci vztahů mezi vyšetřovanými osobami a majetkem.

Katarína Starinská

Parameters estimates for change-point detection problem in AR time series

MFF UK, KPMS, Sokolovská 83, CZ – 186 75 Praha 8
starinska@karlim.mff.cuni.cz

The poster shortly presents score test statistic for detection changes in the parameter values of an autoregressive (AR) time series. This statistic is based on the conditional likelihood function and for evaluating the statistic we need to estimate the parameters of AR. As we already have the likelihood function (resp. conditional likelihood function), we use the maximum likelihood estimates (MLE). To show the asymptotic properties of score test statistic we require the estimates to converge to the true parameters with the rate $O(\sqrt{k^{-1} \log \log k})$. It can be proved that MLE have this rate of convergence, but under some restrictive assumption - we need to assume that all the other parameters are known. We use slightly different estimates with the same rate of convergence as MLE without the need of this assumption.

Acknowledgement: The work was supported by the grants SVV 265 315 and GAUK 586 712/2012.

David Stibůrek

Asymptotická ekvivalence statistik spojitých difúzních procesů pro náhodné časy

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

david.stiburek@gmail.com

Při statistické analýze spojitých difúzních procesů dostáváme mnohdy asymptotické výsledky pro pozorování v deterministických časech, které mají nejčastější volbu $0, \frac{1}{n}, \frac{2}{n}, \dots, 1$ na intervalu $[0, 1]$. Tyto časy však mohou být při praktickém pozorování z části vychýleny.

V této práci přijmeme předpoklad, že je daný proces pozorován v náhodných časech $\hat{t}_k^{(n)} := \sum_{j=1}^k \xi_j^{(n)}$, $k = 1, \dots, n$, $n \in \mathbb{N}$, kde $\xi_j^{(n)}$ jsou nezávislé na procesu, mezi sebou a jsou se středními hodnotami $1/n$ plus ohraničeně asymptoticky menším vychýlením. Potom při rozptylech s ohraničeným řádem do $1/n$ pro tyto přírůstky dostáváme při dostatečné konečné variaci procesu asymptotické výsledky zachovány. V případě symetrie těchto přírůstků kolem své střední hodnoty zůstávají asymptotické výsledky zachovány bez ohledu na velikosti rozptylů $\xi_j^{(n)}$.

Blanka Šedivá

Stabilita optimální volby portfolia Markowitzova modelu

Katedra matematiky Fakulty aplikovaných věd, Západočeská univerzita v Plzni

sediva@kma.zcu.cz

Cílem práce je použít poznatky z oblasti teorie náhodných matic a pomocí tohoto přístupu analyzovat vztahy mezi výnosy akcií obchodovaných na Burze cenných papírů Praha, a.s. v segmentu obchodního systému SPAD. Pomocí metod teorie náhodných matic byla studována stabilita korelační matice výnosů těchto akcií za období od roku 2005 do roku 2012 a vliv změny charakteru korelační matice na změnu optimálního portfolia získaného na základě Markowitzova modelu. Byly testovány hodnoty a vzdálenosti vlastních čísel empirické korelační matice v jednotlivých časových periodách. Další analýzy byly zaměřeny na porovnání empirických výstupů se simulovanými daty obsahujícími různě stupně šumu. Výsledky ukazují, že empirická korelační matice obsahuje značné množství šumu, což může mít, zejména v některých obdobích, velmi výrazný vliv na volbu optimálního portfolia odvozeného z Markowitzovy teorie portfolia.

Pavel Tuček¹, Michaela Tučková¹ a Radoslav Harman²

Optimální návrh měření sigmoidálních funkcí

¹ Univerzita Palackého v Olomouci, Katedra matematické analýzy a aplikací matematiky, Tř. 17. listopadu 12, 771 46, Olomouc, pavel.tucek@upol.cz

² Univerzita Komenského, Katedra aplikovanéj matematiky a štatistiky, Mlynská dolina, 842 48 Bratislava 4, Slovenská republika

pavel.tucek@upol.cz

Příspěvek je věnován specifické aplikaci teorie optimálního navrhování experimentů v nanomateriálovém výzkumu. Konkrétně se zabývá optimalizací procesu měření magnetizace epsilon fáze nanomateriálových sloučenin oxidů železa, jenž disponují velkým aplikačním potenciálem. Výstupem tohoto procesu měření jsou hysterézní smyčky, k jejichž aproximaci jsou užívány známé sigmoidální funkce (Brillouinova nebo Langevinova) s neznámými parametry. Ty nejenže jednoznačně charakterizují zkoumaný nanomateriál, ale zároveň rozhodují o jeho budoucí aplikaci v praxi. Ve snaze naplnit hlavní cíl této tematiky, a získat tedy co nejpřesnější odhady neznámých parametrů obou funkcí, volíme D-optimální kritériální funkci k sestavení vhodného plánu měření. Problematiku nelinearity uvažovaných fyzikálních modelů řešíme vymezením přípustných parametrických prostorů obou funkcí, v nichž následně hledáme maximin eficientní návrh vzhledem k uvažovanému D-optimálnímu kritériu optimality. Tímto postupem získáváme univerzální design pro měření libovolných nanomateriálových sloučenin epsilon fáze oxidů železa.

Literatura

- [1] Atkinson A. C., Donev A. N. a Tobias R. D. Optimum experimental designs, with SAS. Oxford University Press, 2007.
- [2] Darby M. Tables of the Brillouin function and of the related function for the spontaneous magnetization. British Journal of Applied Physics 18, 1415–1417, 1967.
- [3] Müller CH. a Pázman A. Application of necessary and sufficient conditions for maximin efficient designs. Metrika 48, 1–19, 1998.
- [4] Pukelsheim F. Optimal design of experiments. John Wiley and Sons, 1993.

Michaela Tučková¹, Lubomír Kubáček¹ a Pavel Tuček²

Design experimentu pro regresní modely s podmínkami typu I

¹ Univerzita Palackého v Olomouci, Katedra matematické analýzy a aplikací matematiky, Tř. 17. listopadu 12, 771 46, Olomouc, michela.tuckva@upol.cz

² Univerzita Palackého v Olomouci, Katedra geoinformatiky, Tř. Svobody 26, 771 46, Olomouc

michala.tuckova@gmail.com

Regresní modely s podmínkami mají v oblasti regresní analýzy silné zastoupení. Podmínkami typu I myslíme takové funkce parametrů, jejichž proměnnými jsou pouze parametry v modelu se vyskytující. Přestože úvaha hledání lokálně optimálních návrhů měření v regresních modelech s podmínkami stojí na podobném principu jako je tomu v regresních modelech bez podmínek, vyskytují se zde důležité odlišnosti. V příspěvku se zabýváme kritérii lokální A-optimality, C-optimality a D-optimality, pro která odvozujeme konkrétní podoby gradientů a následně demonstrujeme iterační proces hledání lokálně optimálního návrhu v regresním modelu s podmínkami typu I.

Literatura

- [1] Fišerová E., Kubáček L. a Kunderová P. Linear statistical models. Regularity and Singularities. Academia, Praha, 2007.
- [2] Pázman A. Optimal design of nonlinear experiments with parameter constraints. *Metrika* 56, 113–130, 2002.
- [3] Pukelsheim F. Optimal design of experiments. John Wiley and Sons, 1993.
- [4] Wynn H. P. Results in the theory and construction of *D*-optimum experimental designs. *Journal of the Royal Statistical Society, Series B.* 34 (2) 133–147, 1972.

Ondřej Vencálek

Klasifikační metoda k nejbližších sousedů a hloubka dat

Přírodovědecká fakulta Univerzity Palackého v Olomouci

ondrej.vencalek@upol.cz

Cílem příspěvku je představit různé způsoby použití metodologie hloubky dat k modifikaci známé klasifikační metody k nejbližších sousedů.

Metoda k nejbližších sousedů je jednou z nejpoužívanějších metod pro řešení klasifikačního problému. Narozdíl od mnoha dalších metod, jako je lineární či kvadratická diskriminační analýza, není její použití vázáno na úzkou skupinu distribučních funkcí. V první části příspěvku krátce zopakujeme dobré vlastnosti této metody.

V nedávné době se objevily dva různé postupy, jak metodu k nejbližších sousedů modifikovat pomocí tzv. hloubky dat. V příspěvku oba tyto postupy představíme a porovnáme.

Viktor Witkovský

Exaktné testy a konfidenčné oblasti pre parametre normálneho lineárneho modelu s dvomi variančnými komponentami

Ústav merania SAV, Bratislava

witkovsky@savba.sk

V príspevku popíšeme metódu konštrukcie konfidenčných oblastí založených na invertovaní exaktných testov založených na funkcii podielu vierohodností (LRT/RLRT) pre parametre normálneho regresného modelu s dvomi variančnými komponentami

$$Y \sim N_n(X\beta, \sigma^2 W(\lambda)),$$

kde X je známa ($n \times p$) matica, $\beta \in R^p$ je neznámy vektor parametrov a $\sigma^2 W(\lambda) = \sigma^2(I_n + \lambda V)$ je kovariančná matica, pričom V je známa nezáporne definitná matica, ktorá závisí od neznámych parametrov $\sigma^2 > 0$ a $\lambda \geq 0$, pozri tiež [1, 2, 3, 4].

Pod'akovanie : Práca vznikla v spolupráci s J. Volaufovou (School of Public Health, LSU Health Sciences Center, New Orleans, USA) a vďaka podpore grantov APVV-0096-10, VEGA 2/0019/10 a VEGA 2/0038/12.

Literatúra

- [1] Crainiceanu C. M. a Ruppert D. Likelihood ratio tests in linear mixed models with one variance component. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 66, 165–185, 2004.
- [2] Chvosteková M. a Witkovský V. Exact likelihood ratio test for the parameters of the linear regression model with normal errors. *Measurement Science Review*, 9(1), 1–8, 2009.
- [3] Volaufová J. a Witkovský V. On exact inference in linear models with two variance-covariance components. *Tatra Mountains Mathematical Publications*, 51, 2012. Accepted for publication.
- [4] Witkovský V. a Volaufová J.: On exact and approximate simultaneous confidence regions for parameters in normal linear model with two variance components. *LINSTAT 2012 – The International Conference on Trends and Perspectives in Linear Statistical Inference & IWMS 2012 – The 21st International Workshop on Matrices and Statistics*, July 16–20, 2012, Będlewo, Poland.

Markéta Zikmundová¹, Kateřina Staňková Helisová², Viktor Beneš¹

Užití částicového marginálního Metropolisova Hastingsova algoritmu ve stochastické geometrii

¹MFF UK, KPMS, Sokolovská 83, 186 75 Praha 8 – Karlín, ²FEL ČVUT, katedra matematiky, Technická 2, 166 27 Praha 6

zikmundm@karlin.mff.cuni.cz, helisova@math.feld.cvut.cz, benesv@karlin.mff.cuni.cz

Částicový marginální Metropolisův Hastingsův algoritmus (PMMH), viz [1], kombinuje klasické MCMC metody se sekvenčním Monte Carlo. Na základě časoprostorového parametrického modelu daného sjednocením interagujících kroužků zkoumáme použití PMMH k odhadu jeho parametrů.

Uvažujeme náhodné sjednocení $U_{\mathbf{y}}$ disků \mathbf{y} [4], [3] dané exponenciální hustotou

$$p(\mathbf{y}|\mathbf{x}) = c_{\mathbf{x}}^{-1} \exp(\mathbf{x}(A(U_{\mathbf{y}}), L(U_{\mathbf{y}}), \chi(U_{\mathbf{y}})))$$

vzhledem k referenčnímu Poissonovu bodovému procesu disků, kde $c_{\mathbf{x}}$ je normalizační konstanta, $A(U_{\mathbf{y}})$ značí celkovou plochu sjednocení, $L(U_{\mathbf{y}})$ obvod a $\chi(U_{\mathbf{y}})$ Eulerovu–Poincarého charakteristiku. Parametr $\mathbf{x} \in \mathcal{R}^3$ se vyvíjí v čase k podle přechodové hustoty $p_{\theta}(\mathbf{x}_k|\mathbf{x}_{k-1})$, kde θ je (vícerozměrný) pomocný parametr. Kromě parametru \mathbf{x} uvažujeme časovou závislost i pro jednotlivé konfigurace sjednocení $U_{\mathbf{y}}$. Algoritmus pro generování takového procesu je popsán v [5]. V případě, že je parametr θ známý, nebo jsme-li schopni ho odhadnout jinou metodou, můžeme k odhadu vývoje parametru \mathbf{x} použít klasické sekvenční metody (částicový filtr, [5]).

PMMH algoritmus nám dává možnost odhadu v případě, že parametr θ je neznámý. Jedná se vlastně o kombinaci běžného částicového filtru [2] a Metropolisova Hastingsova algoritmu. V každé iteraci Metropolisova Hastingsova algoritmu nejprve navrhne parametr θ a pak použijeme s tímto již známým parametrem částicový filtr k návrhu \mathbf{x} . Hastingsův poměr je potom kombinací návrhového rozdělení parametru θ , jeho apriorního rozdělení a také na věrohodnosti $p(\mathbf{y})$. Dvojice návrhů (θ, \mathbf{x}) je s pravděpodobností odpovídající Hastingsově poměru přijata.

Příspěvek je simulační studie zkoumající odhad PMMH skrze integrální čtvercovou chybu (MISE) a odhad kovarianční funkce a kontaktní distribuční funkce.

Literatura

- [1] Andrieu C., Doucet A. a Holenstein R. Particle Markov chain Monte Carlo methods. *JRSS B* 72(3), 269–342, 2010.
- [2] Doucet A., de Freitas N. a Gordon N. *Sequential Monte Carlo methods in practice*. Springer, New York 2001.
- [3] Møller J. a Helisová K. Power diagrams and interaction process for unions of discs. *Adv. Appl. Prob.* 40, 321–347, 2008.
- [4] Møller J. a Helisová K. Likelihood inference for unions of interacting discs. *Scand. J. Statist.* 37, 365–381, 2010.
- [5] Zikmundová M., Staňková Helisová K., Beneš V. Spatio-temporal model for a random set given by a union of interacting discs. *Methodology and Computing in Applied Probability*, DOI.

Marta Žambochová

Modifikace algoritmu FEKM

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

marta.zambochova@ujep.cz

Príspevek se zabývá popisem navržené modifikace algoritmu FEKM. Algoritmus FEKM (Fast and Exact K-Means) je jednou ze známých variant algoritmu k -průměrů umožňující zpracování velmi rozsáhlých datových souborů.

Hlavní myšlenkou algoritmu je prvotní vytvoření přiměřeně velkého výběrového souboru z původního souboru dat. V rámci tohoto souboru jsou vytvořeny shluky pomocí klasického algoritmu k -průměrů. V jednotlivých iteracích v průběhu shlukování výběrového souboru jsou zaznamenávána všechna centra a k nim předem definované popisné statistiky. Pomocí všech těchto center pak dochází k vytváření cílového shlukování celého souboru při minimálním počtu průchodů celým datovým souborem.

Ve výjimečně příznivém případě stačí pouze jeden průchod celým datovým souborem. V nejhorším možném případě je nutný stejný počet průchodů jako u klasického algoritmu k -průměrů. Malý počet průchodů celým souborem potřebný k provedení celého algoritmu tak, jak deklarují jeho autoři, byl však potvrzen pouze ve výjimečných případech. Počet průchodů silně závisí na prvotním výběrovém vzorku dat. Algoritmus FEKM tvoří vzorek dat náhodným výběrem. Hlavní myšlenkou navrhované modifikace je proto vytvoření vzorku dat nikoliv náhodně, ale za pomoci jistých datových struktur (stromů). Výsledný vzorek pak lépe vypovídá o rozložení dat a následně se sníží potřebný počet průchodů celým datovým souborem.