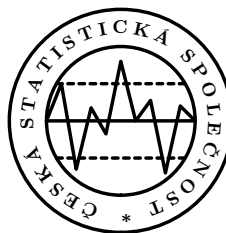
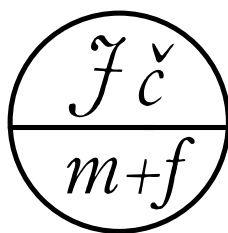
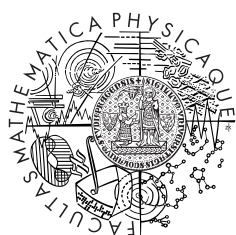


NOVOHRADSKÉ STATISTICKÉ DNY 2012



NOVOHRADSKÉ STATISTICKÉ DNY 2012

PROGRAM

			NEDELE
13.00			Oběd
14.00			
	50 10	Marie Hušková přestávka	Change-point analysis
15.00			
	50 10	Daniela Jarušková přestávka	Detection of changes in mean of vectors with application to hydrology
16.00			
	40 15 15 10	Zuzana Prášková Katarína Starinská Marek Dvořák přestávka	Detekce změn v autoregresních modelech Detekcia zmien v autoregresných postupnostiach so závislými chybami Efficient score test for change detection in vector AR models
17.20			
	20 20 20	Ivana Horová Marie Budíková Jitka Bartošová	Jádrová regrese pro model s korelovanými pozorováními Význam váhové funkce v linearizovatelných regresních modelech Reakce monetární chudoby na změny v modelu spotřební jednotky
18.20		konec	
			PONDĚLÍ
8.30			
	50 10	Ladislav Budík přestávka	Metody odhadu parametrů časových řad průtoků
9.30			
	20 30 10	Marie Forbelská Jan Pícek přestávka	Smíšené regresní modely při sledování obsahu těžkých kovů Odhady návratových hodnot pro srážkovou a teplotní data
10.30			
	25 25 10	Jaromír Antoch Petr Císař přestávka	Poznámka k modelování vertikální atmosferické radioaktivity Design of experiment, application in biology
11.30			
	25	Jiří Dušek	Ekosystémová stanice Mokré louky u Třeboně
12.30			Oběd
14.00			
	50 10	Arnošt Komárek přestávka	Klasifikace na základě longitudinálních dat
15.00			
	30 15 15 10	Josef Tvrdlík Hana Řezanková Marta Žambochová přestávka	Hybridní algoritmus pro optimální shlukování Poznámky k problematice určování počtu shluků Shlukovací algoritmy založené na vzorkování
16.10			
	40 20 10	Dalibor Štys David Legát přestávka	Image information, its representation and multivariate analysis MCMC perfect sampling a analýza obrazu
17.10			
	90	Daniel Hlubinka	Hloubka dat, mnohorozměrné kvantily a jejich použití
18.40		konec	
			ÚTERÝ
8.30			Exkurze na výzkumnou stanici ENKI Třeboň a ekosystémovou stanici Mokré louky u Třeboně

SEZNAM ÚČASTNÍKŮ

Jaromír Antoch	MFF UK	antoch@karlin.mff.cuni.cz
Jitka Bartošová	VŠE JH	bartosov@fm.vse.cz
Ladislav Budík	ČHMU Brno	budik@chmu.cz
Marie Budíková	MU Brno	budikova@math.muni.cz
Petra Cabalková	MU Brno	324014@mail.muni.cz
Petr Císař	JČU	cisar@frov.jcu.cz
Ondřej Černý	MU Brno	211825@mail.muni.cz
Lukáš Drápal	MFF UK	lukas.drapal@gmail.com
Jiří Dušek	CVGZ AV ČR	dusek.j@czechglobe.cz
Marek Dvořák	MFF UK	dvorakm@karlin.mff.cuni.cz
Marie Forbelská	MU Brno	forbel@math.muni.cz
Daniel Hlubinka	MFF UK	hlubinka@karlin.mff.cuni.cz
Terezia Hodasová	MU Brno	211192@mail.muni.cz
Ivana Horová	MU Brno	horova@math.muni.cz
Zdeněk Hrubý	MU Brno	zdenek22.hruby@gmail.com
Marie Hušková	MFF UK	huskova@karlin.mff.cuni.cz
David Legát	MFF UK	legat@karlin.mff.cuni.cz
Marie Leváková	MU Brno	xlevakov@math.muni.cz
Daniela Jarušková	ČVUT	jarus@mat.fsv.cvut.cz
Arnošt Komárek	MFF UK	komarek@karlin.mff.cuni.cz
Milena Kovářová	JČU	kovarova@greentech.cz
Marek Malý	SZÚ	marek.maly@szu.cz
Jan Pícek	TUL	jan.picek@tul.cz
Zuzana Prášková	MFF UK	praskova@karlin.mff.cuni.cz
Hana Řezanková	VŠE	rezanka@vse.cz
Iveta Salingerová	MU Brno	211456@mail.muni.cz
Katarína Starinská	MFF UK	starins@karlin.mff.cuni.cz
Dalibor Štys	JČU	stys@jcu.cz
Nad'a Štysová	JČU	Nada.Stysova@seznam.cz
Josef Tvrdík	OSSU	josef.tvrdik@osu.cz
Michaela Vojtová	MU Brno	272730@mail.muni.cz
Marta Žambochová	UJEP	marta.zambochova@ujep.cz

NOVOHRADSKÉ STATISTICKÉ DNY 2012

Sborník abstraktů

Antoch Jaromír a Daniel Hlubinka	
<i>Poznámka k modelování vertikální atmosferické radioaktivity</i>	2
Bartošová Jitka	
<i>Reakce monetární chudoby na změny v modelu spotřební jednotky v zemích EU</i>	2
Budík Ladislav	
<i>Metody odhadu parametrů časových řad průtoků v libovolném profilu s opravou antropogenního vlivu</i>	2
Budíková Marie a Bohumil Maroš	
<i>Význam váhové funkce v linearizovatelných regresních modelech</i>	3
Císař Petr a kol.	
<i>Design of experiment, application in biology</i>	3
Dušek Jiří a Stanislav Stellner	
<i>Ekosystémová stanice Mokré louky u Třeboně</i>	4
Dvořák Marek	
<i>Efficient score test for change detection in vector autoregressive models</i>	4
Forbelská Marie	
<i>Smíšené regresní modely při sledování obsahu těžkých kovů v sedimentech řeky Moravy</i>	5
Hlubinka Daniel a kol.	
<i>Hlubka dat, mnohorozměrné kvantily a jejich použití v neparametrické klasifikaci</i>	5
Horová Ivana, Jan Koláček a Dagmar Lajdová	
<i>Jádrová regrese pro model s korelovanými pozorováními</i>	5
Hušková Marie	
<i>Change-point analysis</i>	5
Jarušková Daniela	
<i>Detection of changes in mean of vectors with application to hydrology</i>	6
Komárek Arnošt	
<i>Klasifikace na základě sdružených spojitých a diskrétních longitudinálních dat</i>	6
Legát David	
<i>MCMC perfect sampling a analýza obrazu</i>	6
Pícek Jan	
<i>Odhady návratových hodnot pro srážková a teplotní data</i>	7
Prášková Zuzana	
<i>Detekce změn v autoregresních modelech</i>	7
Řezanková Hana	
<i>Poznámky k problematice určování počtu shluků</i>	8
Starinská Katarína	
<i>Detekcia zmien v autoregresných postupnostiach so závislými chybami</i>	8
Štys Dalibor a kol.	
<i>Image information, its representation and multivariate analysis of resulting global variables applied to microscopic images</i>	8
Josef Tvrdík	
<i>Hybridní algoritmus pro optimální shlukování</i>	9
Žambochová Marta	
<i>Shlukovací algoritmy založené na vzorkování</i>	10

Jaromír Antoch a Daniel Hlubinka**Poznámka k modelování vertikální atmosferické radioaktivity**

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

antoch@karlin.mff.cuni.cz

Vertikální profily radioaktivity jsou již delší dobu měřeny pouze na několika málo místech na světě, viz například Bazilevskaya a Svirzhevskaya (1998), CHMI-RADAC, Hatakka et al. (2000) nebo Li et al. (2007). Mezi měřenými veličinami hrají důležitou roli pozorování tzv. beta- a gamma- radioaktivity. Během přednášky ukážeme, jak je možné vertikální profily radioaktivity modelovat pomocí modifikovaných růstových křivek, respektive pomocí nehomogenního Poissonova procesu. Oba přístupy porovnáme.

Literatura

- [1] Antoch J. a Hlubinka D. Data driven modelling of vertical atmospheric radiation. *Journal of Environmental Radioactivity* 102, 1085–1095, 2011.
- [2] Antoch J. a Jarušková D. Testing a homogeneity of stochastic processes. *Kybernetika* 41, 415–430, 2007.
- [3] Bazilevskaya G.A. a Svirzhevskaya A.K. On the stratospheric measurements of cosmic rays. *Space Science Review* 85, 431–521, 1998.
- [4] CHMI-RADAC. Czech Hydrometeorological Institute, Upper Air and Surface Observation Department.
- [5] Hatakka J., Paatero J., Kyro E., Antikainen V. a Viisanen Y. Monitoring of radioactivity in the upper atmosphere with radiosounding. In: *Proceedings of Symposium on Nuclear, Biological and Chemical Threats in the 21st Century (NBC 2000)*, June 13–15, 2000, Helsinki University of Technology, Espo, Finland, 269–270, 2000.
- [6] Li S.W., Li Y.S. a Tsui, K.C. Radioactivity in the atmosphere over Hong-Kong. *Journal of Environmental Radioactivity* 94, 98–106, 2007.

Jitka Bartošová**Reakce monetární chudoby na změny v modelu spotřební jednotky v zemích EU**

KMIH FMJH, Jarošovská 1117, 377 01 Jindřichův Hradec II

bartosov@fm.vse.cz

Chudoba je mnohostranný koncept zabývající se nepřítomností nebo nedostatečným množstvím zdrojů, které jsou obecně považovány za nezbytné pro jednotlivce nebo domácnost v dané společnosti (Atkinson et al, 2002; Saalvedra et al, 2009.). Předložený článek se zabývá otázkou, jak se závěry analýz chudoby v zemích Evropské Unie mění v důsledku změn v nastavení parametrů modelu spotřební jednotky (Consuming unit – CU)? Budeme konkrétně zkoumat reakci vybraných měr monetární chudoby na parametrizaci spotřební jednotky provedenou tak, aby zahrnovala jak definici CU dle metodiky Organizace pro ekonomickou spolupráci a rozvoj (OECD), tak i její modifikaci. Studie citlivosti odezvy měr chudoby na změny parametrů je aplikována na data pocházející z výběrového šetření EU – SILC z roku 2009. Důraz je kladen na grafickou prezentaci výsledků.

Literatura

- [1] Atkinson A.B., Cantillon B., Marlier E. a Nolan B. *Social Indicators. The EU and Social Exclusion*. Oxford University Press, Oxford, 2002.
- [2] Saalvedra B., Nolan B. a Smeeding T. eds. *The Oxford Handbook of Economic Inequality*. Oxford University Press, Oxford, 2009.

Ladislav Budík**Metody odhadu parametrů časových řad průtoků v libovolném profilu s opravou antropogenního vlivu**

ČHMÚ, pobočka Brno, Kroftova 43, 616 67 Brno

budik@chmu.cz

Příspěvek bude věnován výpočtu tzv. katastru M-denních průtoků povrchových vod, což jsou výpočty přibližných průběhů křivek překročení denních průměrných průtoků vodních toků mimo pozorovací body na tocích. Časové řady, z nichž se tyto křivky překročení budou počítat, jsou obecně přirozeně závislé na srážkách, výparu z povodí,

složení půd, vegetaci a geologii podložních hornin. Na základě známých dat se získají odhady parametrů povodí a pak se optimalizací nastavují tak, aby po poskládání průtoků z jednotlivých povodí vyšel v závěrovém měřeném profilu změřený průtok. Tento postup však funguje jen u tzv. neovlivněných (přírozených) průtoků. Ovšem na průtoky, někdy až zásadně, nějakým způsobem působí člověk svou činností (odběry povrchových a podzemních vod, vypouštění již nepotřebné nebo srážkové vody ze zpevněných ploch, manipulace na přehradách, převody vody mezi povodími, atd.). Proto všude měříme v zásadě průtoky ovlivněné a chceme-li průtoky tzv. odovlivněné, musíme si je spočítat (spíše odhadnout). Popis možných postupů s využitím přibližných časových řad o ovlivnění s měsíčním krokem pro odvození křivek překročení tvoří další část příspěvku. Pro převážnou většinu výše zmíněných postupů jsou již napsány funkční programy v Excelu 2003. Závěr příspěvku bude věnován přímo práci s časovými řadami průtoků v měřených profilech na tocích a možnostem odhadu průběhu časových řad průtoků mimo pozorované profily. Zde existuje celá řada problémů, z nichž největší jsou odhady časových posunů jednotlivých průtokových řad, transformace průtoků (snižování vyšších průtoků a zvyšování nižších průtoků dané objemem vody kumulované v toku a částečně v podzemí v blízkém okolí toku). Řešení tohoto problému je nutným předpokladem pro získání tzv. katastru podzemních vod, který by poté bylo možno sestavit. Pak by byly k dispozici lepší odhady chování podzemních vod a využitelnosti jejich zásob. Zdá se, že by mohlo být i odhadnuto chování vegetace vůči zásobě podzemních vod, a to zřejmě nejen podrobněji, ale i přesněji než dosud.

Marie Budíková a Bohumil Maroš

Význam váhové funkce v linearizovatelných regresních modelech

ÚMaS, PříF MU, Kotlářská 2, 611 37 Brno

budikova@math.muni.cz

Příspěvek zkoumá pomocí simulačních metod vlastnosti nelineárního regresního modelu s jednou vysvětlující proměnnou, v němž je možno nelineární regresní funkci převést vhodnou transformací na funkci, která je lineární v parametrech. Při použití transformační funkce však dojde k porušení normality náhodných odchylek a porušení homogenity jejich rozptylů. Tyto problémy lze aspoň částečně eliminovat tím, že se na linearizovanou regresní funkci aplikuje váhová funkce, případně i opakovaně. V praxi se však s váhovou funkcí setkáme jen zřídka a ani v některých profesionálních statistických systémech není implementována. Uživatelé se rovněž vyhýbají odhadu parametrů nelineárního modelu pomocí numerických metod, neboť se obávají problémů s počátečními odhady parametrů či problémů s konvergencí těchto metod. Proto nejčastěji volí pro odhad parametrů linearizovaného regresního modelu obyčejnou metodu nejmenších čtverců, i když může být zatížena značnými nepřesnostmi. V našem příspěvku se zaměřujeme na exponenciální regresní model tvaru $y = \exp\{\beta_0 + \beta_1 x\}$, v němž sledujeme vliv různých faktorů na vhodnost použití váhové funkce. Tento vliv posuzujeme pomocí reziduálního součtu čtverců a po ukončení iteračního procesu aplikace váhové funkce vyjádříme kvalitu výsledného proložení relativním zlepšením reziduálního součtu čtverců. Výsledky regresní analýzy bez váhové funkce a s opakovanou aplikací váhové funkce ukážeme na konkrétních datech popisujících množení bakterií v tekutém prostředí.

Petr Císař, Martina Tesařová, Zuzana Antošová and Oksana Degtjarika

Design of experiment, application in biology

Faculty of Fishery and Protection of Waters, School of Complex Systems, University of South Bohemia, Zámek 136, 373 33 Nové Hradky

cisar@frov.jcu.cz

The optimization of biological experiments, e.g. the optimization of growth condition, has an increasing importance with an increasing number of diverse molecules produced by these methods. This is due to the fact that the optimization makes the experiments efficient in time, reduces resources and costs. The most extensive optimization in the biological community is typically based only on the experience and intuition of the experimenter without any statistical techniques. The OVAT (One Variable At a Time) optimization technique [1] is widely used by the experimenters. The method has many disadvantages but it is the most popular method for the students who were not educated in any statistically based experiments.

The application of the method Design of experiments (DoE) [2] to optimize culture growth conditions of *Escherichia coli* (*E. coli*) will be described in the presentation. Studied factors (temperature, time of harvest, RPM of shaker and start OD) were chosen based on the preliminary knowledge [3], [4] and the maximal production of the studied genetically inserted protein was observed. Determination of the precise amount of the target protein by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS PAGE) followed by the calibration and set of experiments based on the DoE method was provided with the aim to reduce time consuming experiments and to understand to the behavior of the system. Influences of factors on the production of the target protein and interactions between them were evaluated and the optimum was detected, specified and verified by subsequent experiments. The highest yield of the pure protein was achieved by the calculated combination of the factors.

References

- [1] Mandenius C.F. and Brundin A. Bioprocess optimization using design-of-experiments methodology. *Biotechnol. Prog.* 24, 1191–1203, 2008.
- [2] Antony J. Design of Experiments for Engineers and Scientists. Butterworth Heinemann, an imprint of Elsevier, Oxford, 2003.
- [3] Nikerel I.E. et al. Optimizing medium composition for TaqI endonuclease production by recombinant *Escherichia coli* cells using response surface methodology. *Process. Biochem.* 40, 1633–1639, 2005.
- [4] Sivashanmugam A. et al. Practical protocols for production of very high yields of recombinant proteins using *Escherichia coli*. *Protein Sci.* 18, 936–948, 2009.

Jiří Dušek a Stanislav Stellner

Ekosystémová stanice Mokré louky u Třeboně

Centrum výzkumu globální změny AV ČR, v.v.i. Na Sádkách 7, České Budějovice

dusek.j@czechglobe.cz

V současnosti je patrné negativní působení člověka na jednotlivé ekosystémy v globálním měřítku celé planety. Narušování a poškozování ekosystémů ovlivňuje jejich roli v koloběhu látek a energie. Velká pozornost je věnována koloběhu uhlíku a vody v souvislosti se skleníkovým efektem vodní páry, oxidu uhličitého a metanu v atmosféře. Pro získání znalostí týkající se toku uhlíku a vody je budována celosvětová síť monitorovacích stanic v různých druzích ekosystémů (ICOS = Integrated Carbon Observation System). Tato síť využívá pro sledování toků látek metodu vířivé kovariance (eddy covariance).

Mezi důležité ekosystémy v transformaci látek a koloběhu vody bezesporu patří mokřadní ekosystémy. V mokřadním ekosystému zaplavovaných „Mokřých luk u Třeboně“ jsou dlouhodobě sledovány meteorologické parametry od roku 1977 až do současnosti. V roce 2004 zde byla vybudována nová ekosystémová stanice, která je součástí sítě ICOS. V roce 2005 zde byla instalována technologie pro sledování toků uhlíku (oxid uhličitý) a vodní páry metodou vířivé kovariance. V současnosti jsou také měřeny emise oxidu uhličitého a metanu metodou automatických emisních komor.

Získávání velkého množství dat pomocí automatických zařízení klade velké nároky na jejich kvalitní statistické zpracování. Data změřená metodou vířivé kovariance jsou numericky korigována na podmínky měření vycházející z aerodynamických předpokladů metody. Data jsou klasifikována dle kvality. Chybná a nepřesná měření jsou domodelována. Získáváme kontinuální data o čisté výměně látek (CO_2 a H_2O) mezi atmosférou a ekosystémem. Data jsou interpretována v souvislosti s meteorologickými podmínkami na lokalitě (sucho, záplavy apod.)

Tento text je výstupem Centra CzechGlobe, jehož vybudování je v rámci OP VaVpI spolufinancováno z fondů EU a státního rozpočtu ČR (Projekt: CzechGlobe – Centrum pro studium dopadů globální změny klimatu, Reg. č. CZ.1.05/1.1.00/02.0073).

Marek Dvořák

Efficient score test for change detection in vector autoregressive models

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

dvorakm@karlin.mff.cuni.cz

Most of the articles about change point detection are focused on the univariate models. Detecting changes in multivariate series is receiving an increasing attention nowadays and the test statistics are usually based on likelihood ratio, see for instance paper by Qu and Perron (2007). We will discuss the different approach based on the efficient score vector test statistic. Its asymptotic properties under the assumption of no change and results from the simulation study will be presented.

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

References

- [1] Gombay E. Change Detection in Autoregressive Time Series. *Journal of Multivariate Analysis* 99, 451–464, 2008.
- [2] Lütkepohl H. New Introduction to Multiple Time Series Analysis. Springer, Berlin, 2005.
- [3] Qu Z., Perron P. Estimating and Testing Structural Changes in Multivariate Regressions. *Econometrica* 75, 459–502, 2007.

Marie Forbelská

Smíšené regresní modely při sledování obsahu těžkých kovů v sedimentech řeky Moravy

ÚMaS, PřF MU, Kotlářská 2, 611 37 Brno

forbel@math.muni.cz

Pro celou řeku Moravu byl vytvořen jediný lineární regresní model s náhodným posunutím, který v sobě zahrnuje regresní přímky pro jednotlivé lokality a také regresní přímku pro celou řeku Moravu.

Uvažovaný model je speciálním typem lineárního regresního modelu, a to tzv. lineárního regresního modelu s náhodnými i pevnými efekty, také nazývaného lineární regresní model se smíšenými efekty (Linear Mixed Effect Model). Neznámé parametry modelu jsou odhadovány pomocí metody REML (Restricted Maximum Likelihood).

Jde o modifikaci ANCOVA modelu, který se dokáže díky náhodným parametrům vyrovnat jednak s nestejnými rozptyly v jednotlivých lokalitách a také s korelovanými daty.

Daniel Hlubinka, Stanislav Nagy, Ondřej Vencálek, Lukáš Kotík a Miroslav Šiman

Hloubka dat, mnohorozměrné kvantily a jejich použití v neparametrické klasifikaci

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

hlubinka@karlin.mff.cuni.cz

Hloubka dat je populární neparametrickou mnohorozměrnou statistickou metodou. Hloubka je, populárně řečeno, zavedení lineárního uspořádání na mnohorozměrné (až nekonečněrozměrné) náhodné veličiny. Hloubka by měla odrážet určité postavení bodu vůči pravděpodobnostní míře. Jedna z názorných „definic“ používá průměr s ochranou významné osoby, která by měla být ze všech možných směrů (s ohledem na pravděpodobnostní rozdělení) co nejlépe chráněna.

V přednášce si řekneme něco o hloubce a možnosti definovat mnohorozměrné kvantily na základě hloubky. Zmíníme i semiparametrický přístup k mnohorozměrným kvantilům pomocí regresních kvantilů. Ukážeme si, jak hloubka funguje i jako redukce dimenze a lze ji pak úspěšně použít ke klasifikaci pomocí takzvaných D-D diagramů. Předvedeme si i novou metodu k -nejblíže hlubokých sousedů. Diskutovat budeme nejdůležitější případy dvou a třírozměrných dat a také funkcionálních (nekonečněrozměrných dat). Zbude-li čas, povíme si i něco o konzistenci hloubky, jaké problémy se musí řešit a jaké techniky lze použít.

Ivana Horová, Jan Kolářek a Dagmar Lajdová

Jádrová regrese pro model s korelovanými pozorováními

ÚMaS, PřF MU, Kotlářská 2, 611 37 Brno

horova@math.muni.cz

Jádrové odhady regresní funkce patří mezi velmi efektivní neparametrické metody. Tyto odhady závisí na jádře, které má úlohu váhové funkce a na vyhlazovacím parametru, který řídí hladkost odhadu. Volba vyhlazovacího parametru je zásadním problémem ve všech typech jádrových odhadů. Existuje celá řada metod pro volbu vyhlazovacího parametru např. metody křížového ověřování a plug-in metody. Tyto metody však selhávají v případě, že rezidua jsou korelována. Pro tyto případy navrhuje modifikované verze existujících metod. Navržené metody jsou aplikovány na teplotní řady.

Marie Hušková

Change-point analysis

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

huskova@karlin.mff.cuni.cz

Change point analysis concerns procedures on stability of statistical models. Typically one has a sequence of observations Y_1, \dots, Y_n obtained at the ordered time points $t_1 < \dots < t_n$ such that the first m observations follow a certain statistical model and after the m -th observation the model changes and the remaining $n - m$ observations follow another model. The point m is unknown and is called change point. The problem is to detect (test H_0 : no change & H_1 : there is a change) and to identify the location of such a change (estimate m).

The above formulated problem is sometimes called *retrospective*. All observations are available on at the beginning of statistical analysis. There is also a *sequential version* where the observations are arriving sequentially

and after each new observations we have to decide whether the observations obtained so far indicate an instability or a change.

The above formulated problems have many variants. It has attracted many researchers both from theoretical and practical point of view. Applications can be found in meteorology, climatology, hydrology or environmental studies, econometric time series, statistical quality control among others.

At first the basic description of the considered models and statistical procedures will be provided and then the talk will focus on various applications for detection and identification of changes in mostly in regression models or time series.

Daniela Jarušková

Detection of changes in mean of vectors with application to hydrology

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

jarus@mat.fsv.cvut.cz

When we study possible trends in precipitation series the main effect of global warming may not be a decrease or increase of annual means but rather a change of distribution of precipitation in a year. As a consequence behavior of discharges during a calendar year may change as well. If the behavior during a year is described by monthly averages, the goal of statistical inference is to detect a change in mean of twelve-dimensional vectors with dependent components, see Horváth et al (1999) and Jarušková (2010). Often, it seems more natural to describe the behavior during a year by daily averages. However here, we encounter a problem that the number of components of studied vectors is larger than the number of observed vectors. In such a case application of the method of principal components is recommended, for more details see Benko et al (2009) and Berkes et al (2009). In our contribution we would like to present test procedures for detecting changes in vector means and discuss their advantages and disadvantages.

References

- [1] Benko M., Härdle W. a Kneip A. Common functional principal components. *Annals of Statistics* 37, 1–34, 2009.
- [2] Berkes I., Gabrys R., Horvath L. a Kokoszka P. Detecting changes in the mean of functional observations. *JRSS B71*, 927–946, 2009.
- [3] Horváth L., Kokoszka P. a Steinebach J. Testing for changes in multivariate dependent observations with an application to temperature changes. *Journal of Multivariate Analysis* 68, 96–119, 1999.
- [4] Jarušková D. Asymptotic behavior of a test statistic for detection of change in mean of vectors. *JSPI* 140, 616–625, 2010.

Arnošt Komárek

Klasifikace na základě sdružených spojitých a diskrétních longitudinálních dat

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

komarek@karlin.mff.cuni.cz

V medicínské (ale i jiné) praxi nezřídka provádíme na jednotlivých pacientech (či jiných jednotkách) opakovaná měření (často v čase) jistých ukazatelů (krevní tlak, dichotomická indikace jistých obtíží, počet epileptických záchvatů během předcházejícího období dané délky, ...). Tato měření často souvisejí s nepřímou pozorovatelnou charakteristikou pacienta, kterou může být typ diagnózy nebo stádium nemoci. Logicky se tedy zajímáme o metodu, která by byla schopna na základě celkového vývoje jednotlivých ukazatelů klasifikovat pacienty (jednotky) do skupin vykazujících podobný vývoj těchto ukazatelů a vykazujících tedy podobný typ diagnózy nebo stádium nemoci.

V rámci příspěvku pojednáme o metodě umožňující automatickou klasifikaci jednotek do zadaného počtu skupin na základě longitudinální historie vývoje daných ukazatelů, které mohou být jak spojitě tak diskrétní. Metoda je založena na bayesovském odhadu směšového rozšíření zobecněného lineárního smíšeného modelu (GLMM). Kromě odhadu parametrů modelu a následné klasifikaci pojednáme též o možných metodách volby počtu skupin, do kterých se má klasifikace provádět, není-li tento počet předem znám. Výklad bude doplněn ukázkami na reálných datech a stručným popisem souvisejícího programového vybavení.

David Legát**MCMC perfect sampling a analýza obrazu**

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

legat@karlin.mff.cuni.cz

Mnoho statistických procedur, především v Bayesovské statistice, vede na tak složité modely, že je obtížné nebo dokonce nemožné získat formuli pro odhady parametrů. V takovém případě jsou obvykle používány simulační metody, jako je zamítací výběr nebo MCMC (Markov chain Monte Carlo) simulace tak, že charakteristika pravděpodobnostního rozdělení modelu je odvozena ze získaného vzorku. MCMC simulace generují Markovův řetězec s požadovným stacionárním rozdělením, a proto členy vygenerované sekvence, které jsou dostatečně daleko od počátku, mají přibližně potřebné rozdělení. Nicméně s tímto přístupem je spojeno několik zásadních otázek. Například: „Jak dlouhou sekvenci potřebujeme vyprodukovat, abychom byli dostatečně blízko k požadovanému rozdělení?“ Nebo: „Je možné generovat stacionární Markovův řetězec, jehož stacionární rozdělení je předem určené?“ V průběhu příspěvku budou zmíněny základní simulační metody MCMC jako je Gibbsův a Metropolisův-Hastingsův algoritmus. Následně bude popsána metoda, která umožňuje generovat počáteční prvek Markovova řetězce s požadovaným stacionárním rozdělením, tak aby celý řetězec byl stacionární. Postup bude předveden na vzorku z Isingova modelu.

Zuzana Prášková**Detekce změn v autoregresních modelech**

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

praskova@karlin.mff.cuni.cz

V příspěvku bude podán přehled stávajících metod pro detekci změn v autoregresních modelech s konstantními i náhodnými koeficienty. Zabývat se budeme jak retrospektivními, tak sekvenčními monitorovacími metodami a ukážeme některé aplikace.

Jan Píček**Odhady návratových hodnot pro srážková a teplotní data**

FP TUL, KAP, Studentská 2, 461 17 Liberec

jan.picek@tul.cz

Příspěvek se věnuje odhadům návratových hodnot a jejich neurčitosti pro klimatologická data. Zdánlivě jednoduchá statistická úloha, tj. odhad vysokých kvantilů, má však řadu úskalí. Zajímají nás především vysoké hodnoty srážek, teplot apod., ale na druhé straně máme k dispozici obvykle relativně krátkou dobu pozorování a je tedy otázkou, zdali běžně používané modely založené na asymptotickém rozdělení jsou vhodné. Příspěvek proto diskutuje různou volbu modelu a odhadu parametrů a na reálných datech ukazuje vliv této volby na konečný výsledek. Pozornost je věnována také situaci, kdy je v datech přítomen trend.

Důležitou úlohou je také určení intervalů spolehlivosti. Příspěvek ukazuje některé problémy spojené s jejich konstrukcí na srážkových datech pocházející z Libereckého kraje za období 1960–2010. Diskutuje mimo jiné vliv extrémních srážek v srpnu 2010, které měly za následek záplavy spojené s rozsáhlými materiálními škodami a ztrátami na životech.

Literatura

- [1] Hosking J.R.M. L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society, Series B* 52, 105–124, 1990.
- [2] Hosking J.R.M. a Wallis J.R. *Regional Frequency Analysis: An Approach Based on L-Moments*. Cambridge University Press, New York, 1997.
- [3] Kyselý J. a Píček J. Regional growth curves and improved design value estimates of extreme precipitation events in the Czech Republic. *Climate Research* 33, 243–255, 2007.
- [4] Kyselý J., Gaál L., Píček J. a Schindler M. Estimating return periods for the August 2010 heavy precipitation in northern Bohemia (Czech Republic) in present climate and under climate change, 2012. Zasláno do *Journal of Hydrology*.

Hana Řezanková

Poznámky k problematice určování počtu shluků

VŠE, KSTP, nám. W. Churchilla 4, 13067 Praha 3

rezanka@vse.cz

V souvislosti se zvyšujícím se zájmem firem co nejvíce využít existující datové soubory rozličného charakteru, případně se zájmem provádět rozličné výzkumy, jsou stále více aplikovány metody analýzy dat. Jednou ze zkoumaných oblastí je klasifikace objektů. Pokud analytik nemá informace o tom, jak lze objekty klasifikovat (není známo zařazení žádného z objektů, obvykle není znám ani počet skupin), jsou aplikovány metody shlukové analýzy. Jako příklady využití těchto postupů můžeme zmínit segmentaci zákazníků (klientů) nebo klasifikaci dokumentů či webových stránek.

Uvedená oblast analýzy je poměrně složitá z několika důvodů. Jednak existuje velké množství metod a postupů a pokud v datech neexistuje výrazná struktura, výsledné zařazení objektů do skupin se může lišit. Výsledky některých metod navíc závisí na pořadí, v jakém do analýzy vstupují datové vektory charakterizující sledované objekty. Další problematickou oblastí je určování počtu shluků. Za dobu, po kterou jsou metody shlukové analýzy vyvíjeny, bylo pro tento účel navrženo poměrně velké množství koeficientů. Protože je obtížné určit, který z nich je ten nejlepší, programové systémy nabízejí obvykle několik z nich (pokud vůbec nějaké možnosti nabízejí).

V literatuře lze nalézt řadu studií srovnávajících hodnocení kvality shlukování (včetně stanovení vhodného počtu shluků) pomocí různých koeficientů. Studie jsou obvykle dvou typů, podle toho, zda jsou koeficienty určeny pro hodnocení výsledků pevného shlukování v případě, kdy nepředpokládáme překrývající se shluky (každý objekt je jednoznačně přiřazen právě jednomu shluku), nebo pro hodnocení výsledků fuzzy shlukování, kdy je každé dvojici objekt-shluk přiřazen stupeň příslušnosti objektu ke shluku. Protože v druhém případě může být výsledkem rovněž pevné shlukování (objekt je přiřazen ke shluku podle nejvyšší hodnoty stupně příslušnosti), lze využít oba typy koeficientů. Jejich porovnání naznačuje tento příspěvek.

Provedení prezentované studie bylo inspirováno možnostmi programového systému S-PLUS, v němž jsou v rámci procedury pro fuzzy shlukovou analýzu (algoritmus FANNY) implementovány dva základní koeficienty, a to Dunnův (ve variantě základní a normalizované) a průměrná obrysová šířka (average silhouette width). Jsou-li shluky objektů dobře rozlišitelné, pak pomocí normalizovaného Dunnova koeficientu i průměrné obrysově šířky získáme odpovídající počet shluků. Provedené experimenty naznačují, že v případě, kdy shluky dobře rozlišitelné nejsou (z toho důvodu je právě použita fuzzy shluková analýza), je v rámci zkoumaného intervalu počtu shluků nejlepší variantou pro normalizovaný Dunnův koeficient horní hranice zadaného intervalu, u průměrné obrysově šířky pak dolní hranice zadaného intervalu. Takové rady uživateli zřejmě moc užitečné nebudou, což by mohlo být výzvou pro návrh nového koeficientu, který by kombinoval principy stávajících.

Katarína Starinská

Detekcia zmien v autoregresných postupnostiach so závislými chybami

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

starins@karlin.mff.cuni.cz

Autoregresné modely (AR) sú často využívané na popis javov, ktorých súčasné hodnoty závisia na minulých hodnotách. Najjednoduchším prípadom AR je model s nezávislou rovnako rozdelenou (i.i.d.) postupnosťou náhodného šumu. Pre tento prípad sú známe rôzne testové štatistiky na odhalenie zmeny v parametroch týchto modelov a ich vlastnosti. My sa zameriame na detekovanie zmien v parametroch autoregresných modelov v prípade neštandardných náhodných chýb. Zaujímá nás hlavne náhodný šum, ktorý má vlastnosti martingalových diferencií. Ako testovú štatistiku uvažujeme štatistiku založenú na vektore skórov. Za platnosti vhodných predpokladov ukážeme asymptotické vlastnosti skórovej testovej štatistiky. Teoretické výsledky sú podporené krátkou simulačnou štúdiou.

Podakovanie. Práca bola podporená grantmi SVV 265 315 a GAUK 586 712/2012.

Dalibor Štys, Karina Romanova, Anna Zhyrova, Jan Urban, Jindrich Soukup and Petr Císař

Image information, its representation and multivariate analysis of resulting global variables applied to microscopic images

Faculty of Fishery and Protection of Waters, University of South Bohemia, Zámek 136, 373 33 Nové Hradky
stys@jcu.cz

We present here a practical method for classification of images which is based on calculation of the gain by which each image point contributes to the image information (point information gain - PIG). This approach was

originally developed and is successfully used for image enhancement. The method is based in principle on the proposal of Rényi [1] who proposed a difference of (Rényi) entropies as one of good classifiers characterising the information difference between two distributions. The two probability distributions in our approach arise from the frequency of occurrence of points of given color and intensity in the image: one with and another without the point for which the PIG is calculated [3, 2, 5]. In the resulting image we give to each point in each color channel intensity according its relative contribution to the information in a given image. A transformed image is obtained which best fits the a priori chosen type of distribution.

Empirically we found out that by summation of point information gain we may derive two types of global features, the point information gain entropy (PIE) and the point information gain entropy density (PIE/points). In calculation of PIE we sum all information contributed by all points, in calculation of PIE/points we sum only PIG levels obtained in the resulting histogram. These features form a point in the feature space whose dimensionality is given by number of distributions which was chosen for calculation. PIE and/or PIE/points are sufficiently unique identifiers of each image and may be used for their classification. We also propose possible relation of these variables to generalised dimension of the chaotic attractor which generates the observed pattern [4].

On an example of the microscopic images of living cells and polycrystalline polymers we show that set of PIE values may be used for classification of images in the similar way as the Kullback-Leibler divergence is often used [6]. Although the question whether PIE or PIE/points define a metrics in the information space remains open, we have successfully used the multivariate analysis to classify images in the PIE/points coordinate space.

References

- [1] Rényi A.: On measures of information and entropy. Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability 1960. 547–561, 1961.
- [2] Štys D., Vaněk J., Náhlík T., Urban J., Císař P. The cell monolayer trajectory from the system state point of view. *Mol. BioSyst.* 7, 2824–2833, 2011.
- [3] Štys D., Urban J., Vaněk J., Císař P. Analysis of biological time-lapse microscopic experiment from the point of view of the information theory. *Micron.* 41, 478–483, 2010.
- [4] Štys D., Jizba P., Papáček Š., Náhlík T. and Císař P. On measurement of internal variables of complex self-organized systems and their relation to multifractal spectra, 2012, IWSOS 2012, LCNS 7166, 36–47, Kuipers and Heegaard eds. Springer: Heidelberg Dordrecht London New York, ISBN 978-3-642-28582-0.
- [5] Urban J., Vaněk J., Štys D. Preprocessing of microscopy images via Shannon's entropy. *Pattern Recognition and Information Processing*, 283–287, 2009.
- [6] Kullback S. and Leibler R.A. On Information and Sufficiency *Ann. Math. Statist.* 22, 79–86, 1951.

Josef Tvrđík

Hybridní algoritmus pro optimální shlukování

PřF OU, 30. dubna 22, 701 03 Ostrava 1

josef.tvrdik@osu.cz

Optimální nehierarchické shlukování, tj. seskupení objektů do skupin tak, aby si objekty uvnitř skupin byly co nejpodobnější a skupiny co nejméně podobné, je algoritmicky obtížný problém, který je nutno řešit heuristicky [2, 3]. K ohodnocení kvality shlukování se užívají různá kritéria, např. TRW (trace of within scatter matrix), VCR (variance ratio criterion) nebo MC (Marriottt criterion). Pro hledání optimální klasifikace objektů se obvykle užívá algoritmus k-means, který však často nenajde globální extrém. V prezentaci bude popsán nedávno navržený hybridní algoritmus [4], kombinující adaptivní diferenciální evoluci a k-means. Varianty shlukovacích algoritmů byly porovnány na 8 úlohách z databáze [1] a nový hybridní algoritmus se ukázal jako nejúčinnější.

Literatura

- [1] Asuncion A. a Newman D.J. UCI Machine Learning Repository. Irvine, CA: University of California, School of Information and Computer Science, 2007. <http://www.ics.uci.edu/mllearn/MLRepository.html>
- [2] Das S., Abraham A. a Konar A. Automatic clustering using an improved differential evolution algorithm. *IEEE Transactions on Systems Man, and Cybernetics Part A: Systems and Humans* 38, 218–237, 2008.
- [3] Paterlini S. a Krink T. Differential evolution and particle swarm optimisation in partitional clustering. *CSDA* 50, 1220–1247, 2006.
- [4] Tvrđík J. a Krivý I. Differential evolution with competing strategies applied to partitional clustering. *Lecture Notes in Computer Science* 7269. Berlin Heidelberg: Springer-Verlag, 136–144, 2012.

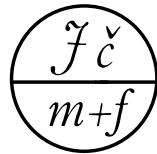
Marta Žambochová

Shlukovací algoritmy založené na vzorkování

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

marta.zambochova@ujep.cz

Príspevek se zabývá popisem a srovnáním vybraných principů algoritmů pro shlukování ve velkých datových souborech. Společným rysem sledovaných algoritmů je řešení problematiky časové náročnosti zpracování tohoto shlukování snížením počtu průchodů celým datovým souborem a dále použitím různých způsobů vzorkování. V některých případech je vzorkování provedeno jednorázovým průchodem datového souboru, při němž jsou konstruovány stromy různých typů (R^* -stromy, CF-stromy). Pomocí takto vzniklých stromů je vytvořen soubor reprezentující strukturu původního datového souboru, ovšem s mnohem menším počtem objektů. Vlastní shlukování již probíhá pouze v rámci tohoto souboru. V jiných případech je vzorek dat vybrán náhodně. Shlukování je provedeno na výběrovém souboru. V dalším postupu těchto algoritmů jsou vytvořené shluky modifikovány v rámci jednoho, či několika málo průchodů původním souborem.



NOVOHRADSKÉ STATISTICKÉ DNY 2012



Vážené kolegyně, vážení kolegové,

ve dnech 3.-5. června 2012 (ne – út) se v prostorách Jihočeské univerzity v Nových Hradech uskuteční *Novohradské statistické dny*. Tato tradiční akce je spoluorganizována Českou statistickou společností (ČStS), Českou matematickou společností při Jednotě českých matematiků a fyziků (ČMS JČMF), Ústavem fyzikální biologie Jihočeské univerzity (ÚFB JU) a grantem „Vzdělávání k posílení konkurenceschopnosti kapacit jihočeského výzkumu a vývoje CZ.1.07/2.3.00/09.0203“.

Novohradské statistické dny budou věnovány především *detekci změn ve statistických modelech s aplikacemi na modelování změn životního prostředí*. Nabídku k přednesení zvaných přednášek přijali prof. Jaromír Antoch, RNDr. Marie Budíková, RNDr. Ladislav Budík, Ing. Petr Císař, doc. Gejza Dohnal, RNDr. Marie Forbelská, doc. Daniel Hlubinka, prof. Ivana Horová, prof. Marie Hušková, prof. Daniela Jarušková, Mgr. Arnošt Komárek, Ing. Štěpán Papáček, doc. Jan Pícek, doc. Zuzana Prášková, doc. Dalibor Štys. S dalšími přednášejícími jednáme. Vedle zvaných přednášek se samozřejmě počítá se sděleními účastníků.

- *Datum a místo konání:* 3.–5. června 2012 (ne – út) v prostorách Akademického a univerzitního centra v Nových Hradech, Zámek 136, 373 33 Nové Hradky.
- *Ubytování:* Ve dvou a třílůžkových pokojích v prostorách Akademického a univerzitního centra v Nových Hradech.
- *Stravování:* Celodenní.
- *Exkurze:* Během NSD 2012 se účastníci seznámí s výzkumným areálem Akademického a univerzitního centra v Nových Hradech a v pondělí odpoledne navštíví Výzkumnou stanici ENKI Třeboň.
- *Sportovně-kulturní program:* Okolí zámku nabízí velmi pěkné prostředí pro milovníky historie i přírody.
- *Technické vybavení:* K dispozici bude tabule, zpětný a datový projektor, notebook a internetové wifi připojení.
- *Abstrakt:* Abstrakt zašlete nejpozději 8. května 2012 jako zdrojový text spolu s pdf souborem.
- *Sborník:* Vybraným příspěvkům bude věnováno speciální číslo Bulletinu České statistické společnosti.
- *Stipendia:* Česká statistická společnost vypsalala několik stipendií umožňujících účast doktorandům. Podrobnosti naleznete na adrese www.statspol.cz.
- *Konferenční poplatek:* 1 600 Kč.
- *Bankovní spojení:* Číslo účtu 200030319/0800 u České spořitelny. Příjemce: Česká matematická společnost při Jednotě českých matematiků a fyziků.
- *Poznámka k placení:* Daňový doklad zašleme po připsání platby na účet a po doručení kopie dokladu o zaplacení, který je třeba k identifikaci platby. Variabilní symbol si určete sami.
- *Další oznámení:* Informace budou uveřejňovány na stránkách ČStS (www.statspol.cz), KPMS MFF UK (www.karlin.mff.cuni.cz/~antoch) a ÚFB JU (www.ufb.jcu.cz/vpk), a distribuovány dle potřeby buď e-mailem nebo klasickou poštou.
- *Adresa pro korespondenci:* MFF UK, KPMS, Novohradské statistické dny, Sokolovská 83, 186 75 Praha 8 – Karlín, tel. 221 913 275; fax. 222 323 316; e-mail antoch@karlin.mff.cuni.cz.

Na setkání se za organizátory těší: J. Antoch, G. Dohnal, D. Hlubinka a N. Štysová

V Praze 7. listopadu 2011