ROBUST 2014
Program a sborník abstraktů

ROBUST 2014 - PROGRAM				
	NEDĚLE	ODPOLEDNE		
	oběd	12.00 - 13.00		oběd bude čekat i na ty, kteří přijedou později
registrace		13.00 - 15.00		
		G. DOHNAL		
J. Antoch		15.00 - 15.05	5	ROBUST 2014 : Zahájení
I. Mizera		15.05 - 16.05	60	Využitie skúsenosti v predikcii: Empirické bayesovské metódy I
	přestávka	16.05 - 16.15		
I. Mizera		16.15 - 17.15	60	Využitie skúsenosti v predikcii: Empirické bayesovské metódy II
	káva	17.15 - 17.35		
		M. MALÝ		
Z. Šulc		17.35 - 17.55	20	Porovnání nových přístupů v oblasti měr podobnosti pro kategoriální data
N. Kaspříková		17.55 - 18.15	20	Některé potíže s klasifikačními modely v praxi
M. Žambochová		18.15 - 18.35	20	Dvoufázový způsob vytváření nekonvexních shluků využívající metody k-průměrů
J. Bartošová		18.35 - 18.55	20	Shlukování prostřednictvím konečných směsí
	večeře	19.00 - 20.00		
	NEDĚLE	VEČER		
		Z. HLÁVKA		
J. Černý		20.15 - 21.15	60	Mathematica, R a SQL

	PONDĚLÍ	DOPOLEDNE		
		P. VOLF		
Z. Pawlas		8.30 - 9.00	30	Statistika Poissonových modelů pro sjednocení kruhů
K. Helisová		9.00 - 9.30	30	Redukce dimenze ve zobecněném Quermass-interakčním procesu
	káva	9.30 - 9.50		
		Z. PAWLAS		
J. Dvořák		9.50 - 10.10	20	Časoprostorové shot-noise Coxovy bodové procesy
A. Koubek		10.10 - 10.30	20	Časoprostorová separovatelnost a ambitové procesy
M. Zikmundová		10.30 - 10.50	20	Bodové procesy úseček
	přestávka	10.50 - 10.55		
		K. HELISOVÁ		
D. Coufal		10.55 - 11.15	20	Jádrové odhady hustot v částicovém filtru
E. Bednáriková		11.15 - 11.35	20	Constructing efficient exact designs of experiments
S. Rosa		11.35 - 11.55	20	Optimal trend resistant experimental designs
	oběd	12.00 - 13.00		
	PONDĚLÍ	ODPOLEDNE		
		E. PELIKÁN		
K. Eben a kol.		16.00 - 17.00	60	Modelování a aproximace kovariancí v asimilaci dat
	káva	17.00 - 17.20		
		K. EBEN		
O. Konár a kol.		17.20 - 17.50	30	Predikce roční spotřeby zemního plynu po ceníkových pásmech
A. Komárek		17.50 - 18.15	25	Regrese s korelovanými intervalově cenzorovanými daty
	přestávka	18.15 - 18.20		
		G. WIMMER		
M. Hladík, M. Černý		18.20 - 18.40	20	Jak počítat odhad rozptylu a t-statistiku nad intervalovými daty
J. Antoch		18.40 - 19.00	20	O užití genetických algoritmů pro výpočet rozptylu nad intervalovými daty
	večeře	19.00 - 20.00		
	PONDĚLÍ	VEČER		
		A. KOMÁREK		
T. Jurczyk		20.15 - 21.15		Data mining a grafické programování

	ÚTERÝ	DOPOLEDNE		
		P. LACHOUT		
M. Friesl		8.30 - 9.00	30	Rovnice na časových škálách a náhodné procesy
Klicnarová		9.00 - 9.30	30	Limitní věty pro slabě závislá náhodná pole
	káva	9.30 - 9. 50		
		V. WITKOVSKÝ		
P. Kříž		9.50 - 10.10	20	Probablity limit identification functions
J. Janák		10.10 - 10.30	20	Statistická inference pro stochastické parciální diferenciální rovnice
K. Kadlec		10.30 - 10.50	20	Convergence of the average cost in the case of the jump diffusions
	přestávka	10.50 - 10.55		
		D. HLUBINKA		
D. Stibůrek		10.55 - 11.15	20	Testování hypotéz parametru driftu u stochastických procesů
P. Veverka		11.15 - 11.35	20	On near-optimal conditions for forward-backward stochastic systems
J. Černý		11.35 - 11.55	20	Kalibrace korelace mezi úrokovými sazbami a časem defaultu
	oběd	12.00 - 13.00		
	ÚTERÝ	ODPOLEDNE		
		J. JURECKOVÁ		
L. Klebanov		16.00 - 17.00	60	Pre-limit theorems and their applications
	káva	17.00 - 17.20		
J.A. Víšek		17.20 - 17.45	25	Diagnostics of the robustified least squares
J. Franc		17.45 - 18.05	20	Computational aspects of robustified mixed LS -- TLS estimator
	přestávka	18.05 - 18.10		
P. Volf		18.10 - 18.35	25	On competing risks and problem of identification
P. Novák		18.35 - 18.55	20	Odhady základního rizika v regresních modelech oprav
	večeře	19.00 - 20.00		
	ÚTERÝ	VEČER		
		G. DOHNAL		
		20.00 - 21.00	60	Beseda o národním parku České Švýcarsko

	STŘEDA	**DOPOLEDNE**		
		I. MIZERA		
J. Jurečková a J. Picek		**8.30 - 9.30**	**60**	**Averaged regression quantiles**
	káva	**9.30 - 9. 50**		
		J. PICEK		
R. Navrátil		**9.50 - 10.10**	**20**	**Pořadové testy v regresi při rušivé heteroskedasticitě**
R. Sabolová a V. Sečkárová		**10.10 - 10.30**	**20**	**I-divergence based statistical inference in exponential family**
Z. Rošťáková		**10.30 - 10.50**	**20**	**Stochastické modelovanie veľkých škôd v poisťovníctve**
	přestávka	**10.50 - 10.55**		
		D. JARUŠKOVÁ		
H. Horáková		**10.55 - 11.15**	**20**	**Odhad změny polohy ročních maxim průtokových řad**
M. Stecenková		**11.15 - 11.35**	**20**	**Klasifikace vzorů v EEG signálu**
R. Zůvala		**11.35 - 11.55**	**20**	**Modelování sesuvu svahu v Halenkovicích pomocí metody kriging**
	oběd	**12.00 - 13.00**		
	STŘEDA	**ODPOLEDNE**		
		M. ŽAMBOCHOVÁ		
	výlet	**13.00 - 18.00**		
	večeře	**18.00 - 19.00**		
	STŘEDA	**VEČER**		
		20.30 - 22.00		**FAB s r.o.**

	ČTVRTEK	**DOPOLEDNE**		
		M. BRABEC		
G. Wimmer		**8.30 - 9.00**	**30**	**A family of transformed Lambert a W*Gamma random variables**
V. Witkovský		**9.00 - 9.30**	**30**	**Poznámky o výpočtových aspektoch testov o pevných a náhodných efektoch**
J. Jakubík		**9.30 - 9.50**	**20**	**Porovnanie metód odhadu fixných efektov v lineárnych zmiešaných modeloch**
	káva	**9.50 - 10.10**		
		S. KATINA		
K. Hron		**10.10 - 10.30**	**20**	**Řídké hlavní bilance**
S. Donevska a kol.		**10.30 - 10.50**	**20**	**Výběr proměnných v kompozičních datech**
K. Fačevicová		**10.50 - 11.10**	**20**	**Statistická analýza kompozičních tabulek**
	přestávka	**11.10 - 11.15**		
		K. HRON		
K. Hrůzová		**11.15 - 11.35**	**20**	**Ekonomická aplikace kompozičního regresního modelu pro odhad rizika**
A. Kalivodová		**11.35 - 11.55**	**20**	**Metoda dílčích nejmenších čtverců pro kompoziční data s aplikací v metabolomice**
P. Kynčlová		**11.55 - 12.15**	**20**	**Aplikace T-prostorů při modelování kompozičních časových řad**
	oběd	**12.15 - 13.15**		
	ČTVRTEK	**ODPOLEDNE**		
		J. BĚLÁČEK		
S. Katina		**16.00 - 17.00**	**60**	**Analýza tvaru a obrazu**
	káva	**17.00 - 17.20**		
		J. KLASCHKA		
M. Kulich		**17.20 - 17.50**	**30**	**Odhadování incidence HIV z průřezových dat**
J. Klaschka		**17.50 - 18.20**	**30**	**O Blakerově konfidenčním intervalu z jiné strany**
	přestávka	**18.20 - 18.25**		
		Z. ROTH		
Z. Hlávka		**18.25 - 18.55**	**30**	**Neparametrické odhady Z-skóre**
J. Běláček		**18.55 - 19.25**	**30**	**O vizualizaci statistických dat II**
	ČTVRTEK	**VEČER**		
		20.00 - 23.59		**závěrečný maškarní ples**

	PÁTEK	DOPOLEDNE		
		Z. FABIÁN		
O. Vencálek		9.00 - 9.30	30	Využití hloubky dat pro klasifikaci -- globální a lokální přístupy
S. Nagy		9.30 - 10.00	30	Konzistencia hĺbky funkcií II
	káva	10.00 - 10.20		
		M. KULICH		
P. Lachout		10.20 - 10.50	30	Poznámka k zápisu náhodných posloupností pomocí polynomů
M. Maciak		10.50 - 11.20	30	Change-point estimation and inference in nonparametric regression
M. Pešta		11.20 - 11.50	30	Trojuholníkové dáta, podmienené najmenšie štvorce, pseudovierohodnosť a kopule
	přestávka	11.50 - 11.55		
		G. DOHNAL		
Z. Fabián		11.55 - 12.40	45	Skórová funkce rozdělení a možné aplikace
J. Antoch		12.40 - 12.45	5	ROBUST 2014 : Ukončení
	oběd	12.45 - 13.45		

Jaromír Antoch

O užití genetických algoritmů pro výpočet rozptylu nad intervalovými daty

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

antoch@karlin.mff.cuni.cz

Cílem přednášky bude ukázat, jak lze použít genetické algoritmy pro výpočet rozptylu (a podobných charakteristik) v případě, kdy máme k dispozici intervalová data, a na jaké problémy přitom uživatel může narazit.

Jitka Bartošová

Shlukování prostřednictvím konečných směsí

KMIH K FMJH, Jarošovská 1117, 377 01 Jindřichův Hradec II

bartosov@fm.vse.cz

TODO

Eva Bednáriková, Radoslav Harman

Constructing efficient exact designs of experiments using a branch-and-bound method

Comenius University in Bratislava, FMPI, Mlynská dolina 842 48 Bratislava 4

eva.bednarikova@fmph.uniba.sk

For a linear regression model, a D-optimal exact experimental design is a sequence of trials that minimizes the volume of a confidence ellipsoid for the unknown parameter. Constructing D-optimal exact experimental designs is a difficult problem of discrete optimization.

To solve this problem, it is possible to use either heuristic methods or exact enumeration methods such as branch-and-bound. The most time consuming computational part of the branch-and-bound algorithms is the repeated use of an optimization method to solve or bound a relaxed convex optimization problem (Welch (1982)). From the statistical point of view, this optimization problem corresponds to the problem of computing a D-optimal linearly constrained approximate experimental design.

In the presentation, we will propose a novel branch-and-bound method that solves the relaxed problems using a multiplicative algorithm for the so-called stratified D-optimality (Harman (2014)). The new approach allows us to develop a rapid branch-and-bound method with a statistically interpretable stopping rule based on the notion of design efficiency.

Literature

[1] Welch W.J. (1982) Branch-and-bound search for experimental designs based on d optimality and other criteria, *Technometrics* **24**, 41 – 48.

[2] Harman R. (2014) Multiplicative methods for computing D-optimal stratified designs of experiments, *Journal of Statistical Planning and Inference* **146**, 82 – 94.

Jaromír Běláček Jaromír

O vizualizaci statistických dat – II

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

jaromir.belacek@lf1.cuni.cz

V intencích stejně pojmenovaného příspěvku na LŠ Robust 2012 se vizualizace dat stává v mnoha případech dokonce nezbytným atributem prezentací výsledků formálních analýz – jmenovitě v případech, kdy statistickou dokumentaci (výsledky analýz) předkládáme většímu laikovi ve statistice, než jsme sami, anebo - jsouce odborníky ve statistice – použijeme při analýze téhož problému kupř. dvě nebo více metod a potřebujeme si vyjasnit KDE, JAK resp. PROČ se výsledky v různých detailech „statisticky významně“ liší (?). Ani pro relativně jednoduché úlohy nám vizualizaci příliš neusnadňuje běžně dostupný statistický SW.

Cílem tohoto příspěvku je demonstrovat na několika příkladech z biomedicínské praxe:

1. „PROČ“ je potřeba zjištěné rozdílné statistické významnosti někdy zdokumentovat i prostřednictvím vhodně zvolených grafických prezentací,

2. „ŽE“ se můžeme setkat již při samotném standardním statistickém zpracování s úlohami, které se bez vizualizace dat de facto vůbec nedají prakticky (a časově) zvládnout.

Za příklad dobře ilustrující první úlohu může sloužit standardní situace u dvouvýběrových testů (nezávislé skupiny pacientů s parodontózou vs kontroly), kdy sledované markery nejsou normálně rozdělené a my jsme dostali věcně ne zcela shodné výsledky na základě několika různých modifikací neparametrických metod (Mann-Whitney, Spearmanovy korelace nebo komparace prostřednictvím výběrových mediánů). Za typický příklad druhé úlohy bude zvolena NLR aplikace v pětiparametrickém modelu tlumeného harmonického kmitání (použitého pro analýzu e-záznamů pádu bérců, loktů a rukou z extenze při výzkumu laterality končetin), kdy vizualizaci dat musíme provést jednak PŘED spuštěním formálního algoritmu (pro specifikaci nosiče/časového rozmezí pro provedení výpočtu), ale také PO výpočtu (z důvodů kontroly kvality proložení dat modelem a pro případný operativní přepočet).

I přes veškerou snahu může „vizualizace dat“ nakonec poukázat na to, že naše dosavadní snažení mohlo být z hlediska jednoznačnosti nebo objektivity závěrů třeba i kontraproduktivní - ale může nám pomoci (mnohem dříve než nám toto řeknou jiní) najít směr resp. cestu (jinou metodu nebo její modifikaci), ve kterém lze naše výsledky přepočítat či opravit tak, aby byly relevantní.

David Coufal

Jádrové odhady hustot v částicovém filtru

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

david.coufal@cs.cas.cz

V přednášce se budeme zabývat problémem neparametrických jádrových odhadů hustot v částicovém filtru. Úloha filtrování spočívá ve stanovení optimálního odhadu nepozorovaných parametrů náhodného procesu na základě pozorovaných dat. Řešení této úlohy vede na stanovení podmíněného rozdělení zájmových parametrů vzhledem k pozorovaným datům. Analyticky lze takové stanovení provést pouze ve speciálních případech. Kanonickým příkladem je lineární Gaussovský proces, pro který je řešením známý Kalmánův filtr.

Vobecném případě se postupuje tak, že se teoretické podmíněné rozdělení aproximuje pomocí empirického rozdělení, které je konstruováno na základě sekvenčního vzorkování. Odhady zájmových parametrů se pak stanoví jako integrální charakteristiky těchto empirických rozdělení. Jako částicový filtr se označuje stochastický algoritmus, který sekvenčně generuje vzorky pro konstrukci empirických rozdělení, a příslušná metodologie jako sekvenční metody Monte Carlo.

Vzorkování v částicovém filtru je prováděno tak, aby bylo teoreticky zaručeno, že se vzrůstajícím počtem vzorků (částic) konstruované empirické míry konvergují k mírám teoretickým. V přednášce ukážeme jak na základě empirických vzorků z částicového filtru konstruovat odhady hustot, které konvergují k hustotám teoretických podmíněných rozdělení. Budeme se rovněž zabývat otázkou vlastností translačního jádra procesu tak, aby uvedená konvergence platila v libovolném čase práce filtru.

Jakub Černý

Kalibrace korelace mezi úrokovými sazbami a časem defaultu

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

jcerny@karlin.mff.cuni.cz

Dle nové bankovní regulace Basel III by banky měly zahrnout do výpočtu kreditní přirážky k tržnímu ocenění (CVA) OTC derivátů také tzv. wrong-way riziko. Wrong-way riziko představuje nepříznivou závislost mezi cenou finančního derivátu, resp. cenou podkladového aktiva, a časem defaultu. Za předpokladu, že propojení mezi sazbou úrokového swapu (IRS) a časem defaultu je reprezentováno gaussovskou kopulou s konstantním korelačním koeficientem, lze toto riziko vyjádřit právě pomocí tohoto korelačního koeficientu. Vzhledem k tomu, že pozorování času defaultu znamená zánik společnosti, nelze tuto korelaci jednoduše odhadnout pomocí pozorovaných dat na rozdíl od intenzity defaultu, která přímo odpovídá sazbě swapu úvěrového selhání (CDS). Na základě dostupných denních cen IRS a CDS sazeb České republiky jsme korelaci odhadli metodou maximální věrohodnosti za předpokladu, že systematický faktor se řídí AR(1) procesem, abychom mohli dekorelovat obě časové řady. Výsledky ukazují, že korelace kalibrovaná na denní data je poměrně vysoká, a proto by wrong-way riziko nemělo být v tomto případě zanedbáváno.

Jakub Černý

Mathematica, R a SQL

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

jcerny@karlin.mff.cuni.cz

Wolfram Mathematica je moderní výpočetní systém vyvíjený více jak 25 let. Mathematica slouží nejen k vědecko-technickým výpočtům na řadě zahraničních univerzit a v předních vědeckých i komerčních institucích (např. CERN, NASA, Intel, Apple,...), ale v současné době představuje i kompletní prostředí pro prezentace, publikace, vizualizace, výpočty i vytváření aplikací. Cílem prezentace je demonstrovat účastníkům možnosti tohoto komplexního softwaru na konkrétních příkladech s hlavním zaměřením na propojení Mathematicy a programu R, práci s velkými daty (propojení Mathematicy a SQL) a novinky týkající se generování procesů.

Sandra Donevska, Peter Filzmoser, Eva Fišerová, Karel Hron

Výběr proměnných v kompozičních datech

UPOL, PřF, KMAaAM a KGI, 17. listopadu 12, 771 46 Olomouc; Department of Statistics and Probability Theory, Vienna University of Technology, Austria

sdonevska@seznam.cz

Častým problémem mnohorozměrné statistické analýzy kompozičních dat je nutnost redukce počtu proměnných (složek) kompozic za účelem zjednodušení struktury dat a snazší interpretace výsledků. Při výběru nadbytečných kompozičních složek, které lze z analýzy vynechat, je nutno obezřetnosti, aby došlo pouze k minimální ztrátě informace o mnohorozměrné datové struktuře při přechodu od kompozice k výsledné podkompozici.

Cílem příspěvku je prezentovat algoritmus pro výběr proměnných v kompozičních datech, který postupně snižuje počet složek kompozic tak, že v každém kroku vynechá tu složku kompozice, která se nejméně podílí na vysvětlení celkové variability kompozičního datového souboru. Přitom je plně využito logratio metodiky kompozičních dat, která adekvátně charakterizuje přirozené vlastnosti tohoto typu pozorování. Výhody použité procedury budou demonstrovány na geochemických datech.

Literatura

[1] Hron, K., Kubáček, L. (2011). Statistical properties of the total variation estimator for compositional data. *Metrika 74*, 221–230.

[2] Hron, K., Filzmoser, P., Donevska, S., Fišerová, E. (2013). Covariance-based variable selection for compositional data. *Mathematical Geosciences 45*, 487–498.

Jiří Dvořák

Časoprostorové shot-noise Coxovy bodové procesy – odhady parametrů a jejich asymptotické vlastnosti

KPMS MFF UK, Sokolovská 83, 186 75 Praha 8; ÚTIA AV ČR, Pod Vodárenskou věží 4, 182 08 Praha 8

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

Představíme velmi flexibilní model nestacionárního časoprostorového shot-noise Coxova bodového procesu s vhodnou strukturou a ukážeme vícekrokovou metodu odhadu parametrů tohoto modelu. Přestože model není časoprostorově separabilní ve smyslu součinového tvaru momentových měr, je možné k odhadu parametrů využít projekcí procesu do prostorové a časové domény.

V prvním kroku jsou odhadnuty parametry popisující nehomogenitu pomocí poissonovské skórové odhadovací rovnice. Ve druhém kroku jsou odhadnuty interakční parametry metodou minimálního kontrastu využívající momentové charakteristiky druhého řádu prostorové a časové projekce procesu.

Tento postup umožňuje odhadnout všechny parametry uvažovaného modelu. Ukazuje se však, že použití metody minimálního kontrastu na časoprostorovou K-funkci s pevným prostorovým dosahem vede v druhém kroku k podstatně přesnějším odhadům některých parametrů. U této vylepšené metody uvedeme, za jakých podmínek je možné dokázat konzistenci a asymptotickou normalitu výsledných odhadů.

Kryštof Eben, Ivan Kasanický˝, Jan Mandel, Pavel Juruš

Modelování a aproximace kovariancí v asimilaci dat

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

eben@cs.cas.cz

Pokrok asimilace dat (aktualizace stavu state space modelu pomocí reálně naměřených hodnot) byl jednou z příčin zpřesnění numerické předpovědi počasí v posledních desetiletích. Asimilační metody se dělí do dvou velkých skupin: ensemblové metody a variační metody. V obou případech je třeba adekvátním způsobem modelovat kovarianční strukturu stavového vektoru a dat. Ensemblové metody, založené obvykle na různých rozšířeních Kalmanova filtru, využívají aproximace kovarianční matice pomocí ensemblu, tj. souboru možných stavů, které nemusí tvořit náhodný výběr a vznikají např. perturbací stavu modelu a vývojem modelu v čase. První takovou metodou byl Ensemblový Kalmanuv filtr (EnKF), kde skutečná kovariancní matice byla aproximována výběrovou kovarianční maticí.

Délka vektoru reprezentujícího stav atmosféry činí řádove několik milionů ci dokonce desítek milionů hodnot. I pri užití nejvýkonnejších superpočítačů tak není možné v praxi počítat s více než stovkami členů ensemblu. V těchto případech, kdy je velikost ensemblu o několik řádů menší než dimenze stavového vektoru, není aproximace pomocí výběrové kovarianční matice příliš vhodná. Proto se používají různé techniky úprav výběrových kovariančních matic (tapering, inflation), které vedou ke zlepšení podmíněnosti, odstranění vychýlení a artefaktních kovariancí.

V přednášce se budeme zabývat úskalími a omezeními pri odhadování skutečné kovarianční matice; další omezení pak vyplývají z poměrně silných předpokladů potřebných pro užití KF. Dále budou představeny odhady kovarianční matice založené na transformaci stavového vektoru do různých spektrálních prostorů a bude zkoumán vztah mezi tvarem kovarianční matice v spektrálním prostoru a stacionaritou náhodného pole reprezentujícího stav. Přesnost některých odhadů bude ilustrována na asimilaci do zjednodušených fyzikálních modelů.

Tato práce byla podpořena Grantovou agenturou České republiky (grant GA13-34856S).

Zdeněk Fabián

Skórová funkce rozdělení a možné aplikace

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

zdenek@cs.cas.cz

Na rozdíl od vektorové Fisherovy skórové funkce je skórová funkce rozdělení skalární funkcí. Funkce se nevztahuje k určitému parametru, ale k jistému středu (typické hodnotě) rozdělení. Protože je skalární, lze ji využít pro robustní odhady parametrů v modelech, které nemají parametr polohy, a při řešení úloh korelační a regresní analýzy.

Kamila Fačevicová, Karel Hron

Statistická analýza kompozičních tabulek

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

kamila.facevicova@gmail.com, hronk@seznam.cz

Příspěvek se zabývá analýzou kompozičních tabulek, které jsou speciálním typem kompozičních dat, popisujícím vztah mezi dvěma faktory. Oproti standardním kontingenčním tabulkám je pro analýzu kompozičních tabulek důležitá pouze relativní struktura dat, což vyžaduje užití tzv. Aitchisonovy geometrie [1]. Cílem statistické analýzy kompozičních tabulek je kvantifikace vztahů mezi faktory v případě, kdy máme k dispozici výběr n kompozičních tabulek. K tomuto účelu slouží rozklad tabulek na jejich nezávislé a interakční části [3, 4], což je možné právě díky vlastnostem Aitchisonovy geometrie. Pro analýzu vztahu mezi faktory jsou pak důležité zejména interakční tabulky, které jsou v případě nezávislosti faktorů rovny neutrálnímu prvku (vzhledem k Aitchisonově geometrii), tedy tabulce, která obsahuje stejné hodnoty. Související statistická analýza interakčních tabulek se značně zjednoduší, pokud použijeme jejich vhodnou transformaci do ortonormálních souřadnic (pomocí tzv. izometrické logratio transformace [2]), ve kterých již můžeme použít standardní metody statistické analýzy včetně statistické inference (např. testování hypotéz). Příspěvek představuje možnou volbu ortonormálních souřadnic pro kompoziční tabulky, které jsou v případě interakčních tabulek snadno interpretovatelné ve smyslu poměrů šancí mezi prvky tabulky. Příspěvek je také doplněn o ilustrační příklad z oblasti fyzické antropologie.

Literatura

[1] Aitchison J (1986) *The statistical analysis of compositional data.* Chapman and Hall, London.

[2] Egozcue JJ, Pawlowsky-Glahn V, Mateu-Figueras G, Barceló-Vidal C (2003) *Isometric logratio transformations for compositional data analysis.* Math Geol 35:279–300.

[3] Egozcue JJ, Díaz-Barrero JL, Pawlowsky-Glahn V (2008) *Compositional analysis of bivariate discrete probabilities.* In Daunis-i-Estadella J, Martín-Fernández JA, eds, Proceedings of CODAWORK'08, The 3rd Compositional Data Analysis Workshop. University of Girona, Spain.

[4] Fačevicová K, Hron K., Todorov V., Guo D. a Templ M. (2013) *Logratio approach to statistical analysis of* 2×2 *compositional tables.* Journal of Applied Statistics, DOI:10.1080/02664763.2013.856871.

Práce je podpořena Operačním programem vzdělávání pro konkurenceschopnost - Evropský sociální fond (projekt CZ.1.07/2.3.00/20.0170 Ministerstva školství mládeže a tělovýchovy České republiky) a grantem PrF_2013_013 - Matematické modely Interní grantové agentury Univerzity Palackého v Olomouci.

Jiří Franc

Computational aspects of robustified mixed LS – TLS estimator

FJFI ČVUT, Trojanova 12, 120 00 Praha 2

jiri.franc@fjfi.cvut.cz

Classical robust regression estimators, such as Least Trimmed Squares (LTS), are not consistent when both independent and some dependent variables are considered to be measured with a random error. One way how to cope with this problem is to use the robustified version of Mixed Least Squares - Total Least Squares (LS-TLS). Mixed Least Trimmed Squares - Total Least Trimmed Squares (LTS-TLTS) based on trimming and mixed Least Weighted Squares - Total Least Weighted Squares (LWS-TLWS) based on the idea of downweighting the influential points, are proposed. The existence and uniqueness of the solution, breakdown point, and another properties of these estimators are discussed. Different approaches of calculation, such as Branch-and-Bound algorithm, elemental concentration algorithm, and Borders Scanning Algorithm, are described and their performances are shown on sets of benchmark instances.

Michal Friesl

Rovnice na časových škálách a náhodné procesy

KM FAV ZČU, Univerzitní 22, pp 314, 306 14 Plzeň

friesl@kma.zcu.cz

Kolegové specializující se na matematickou analýzu se v poslední době zabývají některými typy dynamických rovnic (sjednocující pohled zahrnující prostřednictvím obecné časové škály jak diferenciální rovnice ve spojitém čase, tak diferenční v čase diskrétním) na diskrétním prostoru a zkoumají jejich vlastnosti. Jelikož ve speciálních případech lze řešení některých rovnic chápat jako marginální rozdělení markovského řetězce, je otázkou, zda některé výsledky lze přenést či dovodit ze známých vlastností odpovídajícího náhodného procesu.

Kateřina Helisová[1], Jakub Staněk[2]

Redukce dimenze ve zobecněném Quermass-interakčním procesu

[1]FE ČVUT, KM, Technická 2, 166 27 Praha 6; [2]MFF UK KDM, Sokolovská 83, 186 75 Praha 8

helisova@math.feld.cvut.cz

Uvažujme náhodnou množinu $\mathbf{X}$ danou sjednocením kruhů se středy náhodně rozmístěnými v omezeném okně $S \subset \mathbb{R}^2$ a náhodnými omezenými poloměry. Hustota každé konfigurace $\mathbf{x} = (x_1, ..., x_n)$ kruhů $x_1, ..., x_n$ vzhledem k pravděpodobnostní míře vhodně zvoleného Booleovského modelu (tj. procesu kruhů bez interakcí) je tvaru

$$f_\theta(\mathbf{x}) = c_\theta^{-1} \exp\{\theta \cdot T(U_\mathbf{x})\}, \tag{1}$$

kde $T(U_\mathbf{x})$ je m-dimenzionální vektor geometrických charakteristik (např. plocha, obvod, počet spojitých komponent atd.) sjednocení $U_\mathbf{x}$ kruhů z konfigurace $\mathbf{x}$, $\theta = (\theta_1, \ldots, \theta_m)$ je vektor parametrů, $\cdot$ značí skalární součin a c_θ je normalizační konstanta.

Cílem je zredukovat m-dimenzionální vektor $T(U_{\mathbf{x}})$ na p-dimenzionální vektor $C(U_{\mathbf{x}})$, $p < m$, a tedy v důsledku i hustotu (1) na

$$f_\varphi(\mathbf{x}) = c_\varphi^{-1} \exp\{\varphi \cdot C(U_{\mathbf{x}})\},$$

$\varphi = (\varphi_1, \ldots, \varphi_p)$, popisující dostatečně přesně náhodnou množinu $\mathbf{X}$ (viz [3]). Hlavní motivací k této redukci je zrychlení procedur odhadu parametrů, které jsou při větší dimenzi znatelně časově náročné (např. metoda maximální věrohodnosti využívající MCMC simulací, viz [1]). Příspěvek se zabývá touto redukcí metodou hlavních komponent (viz [2]).

Literatura

[1] Møller J., Helisová K. (2010): Likelihood inference for unions of interacting discs. *Scandinavian Journal of Statistics***37**(3), 365-381.

[2] Rencher A.C. (2002): Methods of Multivariate Analysis, 2nd edn. Wiley & Sons, New York.

[3] Staňková Helisová K., Staněk J. (2013): Dimension reduction in extended Quermass-interaction process. *Methodology and Computing in Applied Probability*, DOI 10.1007/s11009-013-9343-x.

Podporováno granty GA ČR P201/10/0472 a GA ČR 13-05466P.

Milan Hladík[1,2], **Michal Černý**[2]

Jak počítat odhad rozptylu a t-statistiku nad intervalovými daty

[1]MFF UK, KAM, Malostranské nám. 25, 118 00 Praha; [2]VŠE FIS, nám. W. Churchilla 4, 130 67 Praha 3

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.

Nechť $x_1, \ldots, x_n$ je výběr z $N(\mu, \sigma^2)$. Data $x_1, \ldots, x_n$ nejsou pozorovatelná; pozorovatelné jsou hodnoty $\underline{x}_1 \leqslant \overline{x}_1; \ldots; \underline{x}_n \leqslant \overline{x}_n$, o nichž víme, že platí $\underline{x}_1 \leqslant x_1 \leqslant \overline{x}_1, \ldots, \underline{x}_n \leqslant x_n \leqslant \overline{x}_n$. Žádnou jinou informaci o vztahu $\underline{x}_i, x_i, \overline{x}_i$ nemáme. *Statistikou* S rozumíme libovolnou funkci dat $x_1, \ldots, x_n$. Zajímá nás, jakých hodnot může statistika $S(x_1, \ldots, x_n)$ nabývat, známe-li intervalová data $\underline{x}_1, \overline{x}_1; \ldots; \underline{x}_n, \overline{x}_n$. Přesněji řečeno, zajímají nás hodnoty $\underline{S} = \inf\{S(\xi_1, \ldots, \xi_n) : (\forall i = 1, \ldots, n)\ \underline{x}_i \leqslant \xi_i \leqslant \overline{x}_i\}$ a $\overline{S} = \sup\{S(\xi_1, \ldots, \xi_n) : (\forall i = 1, \ldots, n)\ \underline{x}_i \leqslant \xi_i \leqslant \overline{x}_i\}$.

V případě některých statistik je odpověď snadná. Je-li například $S = \widehat{\mu} = \frac{1}{n}\sum_{i=1}^n x_i$, okamžitě zjistíme, že $\underline{S} = \frac{1}{n}\sum_{i=1}^n \underline{x}_i$ a $\overline{S} = \frac{1}{n}\sum_{i=1}^n \overline{x}_i$. Jiným příkladem je odhad rozptylu, je-li μ známé: položíme-li $S = \frac{1}{n}\sum_{i=1}^n (x_i - \mu)^2$, pak $\underline{S} = \frac{1}{n}\sum_{i=1}^n \min\{(\underline{x}_i - \mu)^2, (\overline{x}_i - \mu)^2\}$ a $\overline{S} = \frac{1}{n}\sum_{i=1}^n \max\{(\underline{x}_i - \mu)^2, (\overline{x}_i - \mu)^2\}$.

Situace je ovšem obtížnější v případě odhadu rozptylu, je-li μ neznámé. Položíme-li $S = \widehat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^n (x_i - \frac{1}{n}\sum_{j=1}^n x_j)^2$, lze ukázat, že hodnotu $\underline{S}$ lze spočítat redukcí na konvexní kvadratické programování. To je efektivní algoritmus (= algoritmus pracující v polynomiálním čase). Na druhou stranu ovšem platí, že výpočet $\overline{S}$ je **NP**-těžký problém, a tedy nelze čekat, že by se našla podstatně lepší metoda než výpočet podle vztahu $\overline{S} = \max_{s\in\{0,1\}^n} \frac{1}{n-1}\sum_{i=1}^n [\underline{x}_i + s_i(\overline{x}_i - \underline{x}_i) - \frac{1}{n}\sum_{j=1}^n (\underline{x}_j + s_j(\overline{x}_j - \underline{x}_j))]^2$, který vyžaduje prozkoumání všech 2^n vrcholů krychle $\{0,1\}^n$. Již pro $n = 100$ je takový úkol výpočetně nezvládnutelný.

Ukážeme, že analogická situace platí pro t-statistiku. Položíme-li $t = \sqrt{n} \cdot \frac{\widehat{\mu} - \mu_0}{\widehat{\sigma}}$, kde μ_0 je libovolná konstanta, platí, že hodnotu $\overline{t}$ lze spočítat v polynomiálním čase, zatímco výpočet hodnoty $\underline{t}$ je **NP**-těžký problém.

Platí dokonce více: nejen, že výpočet *přesných* hodnot $\overline{\widehat{\sigma}^2}$ a $\underline{t}$ je **NP**-těžký problém, dokonce i jejich přibližný výpočet je **NP**-těžký, a to ať povolíme *libovolnou* absolutní chybu. (Spočítat hodnotu ξ s absolutní chybou Δ znamená spočítat libovolné číslo ξ^* splňující $|\xi - \xi^*| \leqslant \Delta$.)

Ukážeme, že pro výpočet hodnot $\overline{\widehat{\sigma}^2}$ a $\underline{t}$ existují pseudopolynomiální algoritmy. To je pozitivní výsledek, který umožňuje efektivně vyčíslit hodnoty $\overline{\widehat{\sigma}^2}$ a $\underline{t}$ alespoň v některých speciálních případech. Jsou-li například všechny hodnoty $\underline{x}_1, \overline{x}_1; \ldots; \underline{x}_n, \overline{x}_n$ celočíselné a nevyskytují-li se mezi nimi „příliš“ velká čísla, lze obě hodnoty $\overline{\widehat{\sigma}^2}$ a $\underline{t}$ efektivně spočítat i při velkém n.

Literatura

[1] Chernozhukov V., Hong H., Tamer E. Estimation and confidence regions for parameter sets in econometric models. *Econometrica* 75 (2007) 1243 – 1284.

[2] Černý M., Antoch J., Hladík M. On the possibilistic approach to linear regression models involving uncertain, indeterminate or interval data. *Information Sciences* 244 (2013) 26 – 47.

[3] Černý M., Hladík M. Complexity of computation and approximation of the t-ratio over one-dimensional interval data. Submitted.

[4] Ferson S., Ginzburg L., Kreinovich V. Longpré L., Aviles M. Exact bounds on finite populations of interval data. *Reliable Computing* 11 (2005) 207 – 233.

[5] Gladysz B., Kasperski A. Computing mean absolute deviation under uncertainty. *Applied Soft Computing* 10 (2010) 361 – 366.

[6] Horowitz J.L., Manski C.F. Identification and estimation of statistical functionals using incomplete data. *Journal of Econometrics* 132 (2006) 445 – 459.

[7] Horowitz J.L., Manski C.F., Ponomareva C.F., Stoye J. Computation of bounds on population parameters when the data are incomplete. *Reliable Computing* 9 (2003) 419 – 440.

[8] Kreinovich V. Why intervals? A simple limit theorem that is similar to limit theorems from statistics. *Reliable Computing* 1 (1995) 33 – 40.

[9] Kreinovich V., Longpré L., Patangay P., Ferson S., Ginzburg L. Outlier detection under interval uncertainty: algorithmic solvability and computational complexity. *Reliable Computing* 11 (2005) 59 – 76.

[10] Stoye J. Partial identification of spread parameters. *Quantitative Economics* 1 (2010) 323 – 357.

[11] Vavasis S.A. *Nonlinear Optimization: Complexity Issues.* Oxford University Press, 1991.

Práce byla podpořena granty GAČR P403/12/1947 a P402/13-10660S.

Zdeněk Hlávka

Neparametrické odhady Z-skóre

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

hlavka@karlin.mff.cuni.cz

V příspěvku budou popsány jednoduché parametrické i neparametrické metody konstrukce Z-skóre z naměřených referenčních dat. Hlavním cílem je odvození co nejjednoduššího, ale přitom dostatečně flexibilního a prakticky použitelného postupu. Použité metody jsou demonstrovány na skutečných datech.

Literatura

[1] Šumník, Matysková, Hlávka, Durdilová, Souček & Zemková (2013). Reference data for jumping mechanography in healthy children and adolescents aged 6–18 years. *Journal of musculoskeletal & neuronal interactions* **13**, 259 – 273.

Hana Horáková, Daniela Jarušková

Odhad změny polohy ročních maxim průtokových řad

KM FSv ČVUT, Thákurova 7, 166 29 Praha 6

horakovah@mat.fsv.cvut.cz, jarus@mat.fsv.cvut.cz

Cílem statistického výzkumu je zjistit, zda dochází ke změně chování "ročního chodu" průtokových řad. Pro určitý rok je "roční chod" reprezentován hydrogramem průměrných denních průtoků. Pomocí předchozí statistické analýzy (testování hypotéz) jsme prokázali změnu "ročního chodu" pro 6 z 18 analyzovaných řad.

Jedním z podstatných rysů charakterizující změnu chování je posun jarní kulminace tj. průtoků, které jsou ovlivněny jarním táním. Statistický problém jsme řešili jako dvouvýběrový problém, kde jeden soubor tvoří hydrogramy před rokem 1997 a druhý po roce 1997. Cílem bylo odhadnout posun mezi ročním maximem v prvním a druhém výběru. Pro konstrukci bodového a intervalového odhadu jsme použili dva modely. V prvním modelu jsme předpokládali, že střední hodnota "ročního cyklu" v prvním i druhém období se dá vyjádřit jako jednoduchá cyklická funkce. Pomocí Δ-metody jsme nalezli interval spolehlivosti pro rozdíl maxim obou funkcí. V druhém modelu jsme předpokládali, že v každém roce může nastávat maximum v jiném čase $a_k + \varepsilon_j$ ($k = 1$ odpovídá 1. výběru, $k = 2$ odpovídá 2. výběru) a ε_j jsou náhodné veličiny takové, že $E\,\varepsilon_j = 0$. Pro každý rok jsme maximum $a_k + \varepsilon_j$ odhadli jádrovým odhadem. Interval spolehlivosti pro $a_1 - a_2$ jsme vytvořili pomocí bootstrapu.

Literatura

[1] Jarušková D. Pravděpodobnost a matematická statistika. Praha: Česká technika – nakladatelství ČVUT, 2006.
[2] Anderson T,W. An introduction to multivariate statistical analysis. New York: Wiley & Sons, 1971.
[3] Anderson T,W. The statistical analysis of time series. New York: John Wiley & Sons. 1994.
[4] Prášková Z. Metoda bootstrap. In ROBUST 2004, (J. Antoch and G. Dohnal, eds.), JČMF, Praha, 299 – 314.
[5] Gasser T., Müller H.G. Estimating regression functions and their derivatives by the kernel method. Scandinavian J. Statistics 11, 171 – 185, 1984.
[6] Müller H.G. Kernel Estimators of Zeros and of Locations and Size of Extrema of Regression Functions. Scand. J. Statistics 12, 221 – 232, 1985.

Tato práce byla podpořena grantem Studentské grantové soutěže ČVUT v Praze č. SGS13/005/OHK1/1T/11 "Stochastické metody pro detekci změn v chování hydrogramů".

Karel Hron

Řídké hlavní bilance

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

hronk@seznam.cz

Statistická analýza kompozičních dat pracuje s pozorováními, kde je relevantní informace obsažena pouze v podílech mezi proměnnými, nikoli v původních (absolutních) hodnotách proměnných [1, 4, 5]. V příspěvku se zaměříme na vysoce dimenzionální kompoziční data (ve smyslu stovek či tisíců proměnných), jak se s nimi setkáváme v chemometrii (při hmotnostní spektrometrii), proteomice či genomice. Cílem je provést redukci dimenze uvedených dat tak, aby byly nově získané proměnné (odpovídající hlavním směrům) co nejsnáze interpretovatelné. Ukázalo se, že při nižších dimenzích dat je pro tento účel vhodné použít koncept tzv. hlavních bilancí [6], tedy speciálně volených ortonormálních souřadnic vzhledem k Aitchisonově geometrii kompozičních dat. V tomto případě navrhujeme užití řídkých hlavních komponent (sparse principal components) pro konstrukci hlavních směrů, tzv. řídkých hlavních bilancí [3]. Tyto jsou řídké (obsahují mnoho nul), tvoří ortonormální bázi výběrového prostoru kompozičních dat a efektivně redukují dimenzi dat, je tedy možná jejich aplikace i pro vysoce dimenzionální kompozice.

Literatura

[1] J. Aitchison (1986). *The Statistical Analysis of Compositional Data.* Chapman & Hall, London.
[2] J. J. Egozcue, V. Pawlowsky-Glahn (2005). Groups of parts and their balances in compositional data analysis. *Mathematical Geology*, 37(7), 795–828.
[3] P. Filzmoser, C. Mert, K. Hron (2013). Sparse principal balances. *Statistical Modelling*, přijato k tisku.
[4] K. Hron (2010). Elementy statistické analýzy kompozičních dat. *Informační Bulletin ČStS*, 21(3), 41–48.
[5] V. Pawlowsky-Glahn, A. Buccianti, eds. (2011). *Compositional Data Analysis: Theory and Applications.* Wiley, Chichester.
[6] V. Pawlowsky-Glahn, J.J. Egozcue, R. Tolosana-Delgado (2011). Principal balances. *Sborník konference CoDaWork 2011*, Sant Feliu de Guíxols, Španělsko.

Klára Hrůzová, Karel Hron

Ekonomická aplikace kompozičního regresního modelu pro odhad rizika

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

klara.hruzova@gmail.com

V praktických ekonomických úlohách se často setkáváme s nutností analýzy závislosti mezi dvěma proměnnými vyjádřenými v procentech. Například lze takto zmínit vztah mezi mírou nezaměstnanosti a procentem vysokoškolsky vzdělané populace ve sledovaných zemích. Podobná situace se vyskytuje i při biologických experimentech - na základě zvyšování koncetrace toxické látky se měří procento zasažených živých organismů (tzv. analýza odpovědi na dávku, dose-response analysis). Pro kvantifikaci tohoto vztahu byla zavedena metodika založená na kompozičním charakteru obou proměnných. Konkrétně se jedná o kompoziční regresní model, který ústí v aplikaci standardní regrese pro transformované proměnné (s využitím izometrické logratio transformace) a následnou reprezentaci (a interpretaci) výsledků v původním výběrovém prostoru. Díky uvedeným vlastnostem tohoto modelu je tak možné též využít známé postupy pro příslušnou statistickou inferenci (intervaly spolehlivosti, testování hypotéz) včetně případného využití specifických postupů analýzy odpovědi na dávku. Příspěvek se zabývá interpretací kompozičního regresního modelu při aplikaci na vybrané ekonomické ukazatele.

Literatura

[1] Aitchison, J. (1986). *The Statistical Analysis of Compositional Data*. Chapman & Hall, London.
[2] Egozcue, J. J., Daunis-i-Estadella, J., Pawlowsky-Glahn, V., Hron, K. and Filzmoser, P.(2011). *Simplicial regression. The normal model*. Journal of Applied Probability and Statistics, 6, 87–108.
[3] Egozcue, J. J., Pawlowsky-Glahn, V., Mateu-Figueras, G., Barceló-Vidal, C. (2003). *Isometric logratio transformations for compositional data analysis*. Mathematical Geology 35, 279-300.
[4] Pawlowsky-Glahn, V. and Buccianti, A. (2011). *Compositional Data Analysis: Theory and Applications*. Wiley, Chichester.

Autoři by rádi poděkovali podpoře Operačního programu vzdělávání pro konkurenceschopnost - Evropský sociální fond (projekt CZ.1.07/2.3.00/20.0170 Ministerstva školství mládeže a tělovýchovy České republiky) a internímu grantu Univerzity Palackého v Olomouci PrF_2013_013 Matematické modely.

Jozef Jakubík

Porovnanie metód odhadu fixných efektov vo vysokodimenzionálnych lineárnych zmiešaných modeloch

Ústav merania SAV, Bratislava

jozef.jakubik.jefo@gmail.com

Predstavíme si dve metódy na odhad fixných efektov vo vysokodimenzionálnych lineárnych zmiešaných modeloch z článkov (Rohart2012) a (Schelldorfer2011). Obe metódy sú založené na ℓ_1 penalizácii funkcie maximálnej vierohodnosti, ale používajú rôzne algoritmy. Oboznámime sa s teoretickými predpokladmi daných metód. Porovnáme si ich numerické vlastnosti v rôznych prípadoch.

Lineárne zmiešané modely sú populárne v chovateľstve, zdravotníctve alebo genetike, pretože umožňujú odhadnúť nielen vplyv daného znaku na jednotlivca v skúmanej vzorke, ale na základe skúmanej vzorky aj varianciu daného znaku v celej populácii.

Z daných oblastí taktiež vychádza požiadavka na analýzu vysokodimenzionálnych dát (dáta s väčším počtom znakov ako máme pozorovaní). Spomínané práce sa snažia nájsť podmnožinu množiny efektov, ktorá stále dobre popisuje skúmanú veličinu ale zároveň sa model založený na danej podmnožine jednoduchšie interpretuje. Takýto model sa zvyčajne lepšie správa pri predikcii.

Kľúčové slová: vysokodimenzionálne dáta, lineárny zmiešaný model, ℓ_1 penalizácia

Literatúra

[1] Rohart F., San-Cristobal M., Laurent B. Selection of fixed effects in high dimensional Linear Mixed Models using a multicycle ECM algorithm.
http://florian.rohart.free.fr/Free/Publications_Fr_files/mixed_model_selection_Rohart_et_al_sept2013.pdf
[2] Schelldorfer J., Bühlmann P., van De Geer S. Estimation for high-dimensional linear mixed-effects models using ℓ_1-penalization. Scandinavian Journal of Statistics **2**, 197 – 214, 2011.

Josef Janák

Statistická inference pro stochastické parciální diferenciální rovnice

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

josefjanak@seznam.cz

Uvažujme následující semilineární stochastickou rovnici

$$dX_t = (AX_t + \theta F(X_t))\,dt + \Phi\,dB_t^H, \quad X_0 = 0, \quad t \geq 0,$$

kde $(X_t,\, t \geq 0)$ je V-hodnotový stochastický proces, $(V, \|\cdot\|, \langle\cdot,\cdot\rangle)$ je separabilní Hilbertův prostor, $(B_t^H,\, t \geq 0)$ je standardní cylindrický frakcionální Brownův pohyb s Hurstovým parametrem $H \in (0, 1/2)$, $\Phi \in \mathcal{L}(V)$, $A : \mathrm{Dom}(A) \to V$, $\mathrm{Dom}(A) \subset V$ a A je infinitezimální generátor silně spojité semigrupy $(S(t),\, t \geq 0)$ na V, $F : V \to V$ je nelineární funkce.

Na základě pozorování trajektorie procesu $X^T = \{X_t,\, 0 \leq t \leq T\}$ pořídíme odhad parametru $\theta \in \mathbb{R}$ metodou maximální věrohodnosti a pokusíme se dokázat jeho konzistenci.

Literatura

[1] Duncan T. E., Maslowski B., Pasik-Duncan B. (2009): Semilinear stochastic equations in a Hilbert space with a fractional Brownian motion. *SIAM J. Math. Anal.* **40**(6), 2286–2315.

[2] Kleptsyna M. L., Le Breton A. (2002): Statistical analysis of the fractional Ornstein-Uhlenbeck type process. *Statist. Inference Stoch. Process.* **5**, 229–248.

[3] Tudor C. A., Viens F. G. (2007): Statistical aspects of the fractional stochastic calculus. *The Annals of Statistics* **35**(3), 1183–1212.

Příspěvek vznikl za podpory grantu GA ČR P201/10/0752 a grantu SVV 265 315.

Tomáš Jurczyk

Data mining a grafické programování

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

tomas.jurczyk@statsoft.cz

S rostoucím objemem dat a vývojem výpočetní techniky přišly v analýze dat ke slovu metody data miningu. Data mining jako metodologie pro získávání netriviální skryté informace v rozsáhlých datových souborech se rozšířil a je využíván v mnoha odvětvích – marketing, bankovnictví, genetika, telekomunikace, atd. Program *STATISTICA* je jedním ze softwarů, který implementoval tyto metody a který začal být pro úlohy data miningu využíván, ukážeme si tedy praktické příklady využití této metodologie u některých zákazníků. Dále si také ukážeme grafické prostředí, které se běžně v programech tohoto typu využívá. Zde si můžete pomocí uzlů a jejich napojení nadefinovat celou svou analýzu krok za krokem.

Jana Jurečková[1], Jan Picek[2]

Averaged regression quantiles

[1]MFF UK, KPMS, Sokolovská 83, 186 75 Praha 8; [2] FPHP TUL, KAP, Studentská 2, 461 17 Liberec

jurecko@karlin.mff.cuni.cz, jan.picek@tul.cz

We show that the weighted averaged regression α-quantile in the linear regression model, with regressor components as weights, is monotone in $\alpha \in (0, 1)$, and is asymptotically equivalent to the α-quantile of the location model. This relation remains true under the local heteroscedasticity of the model errors. Among applications, we shall mention the scale statistics based on the averaged regression quantiles, useful for studentization and standardization in linear model. We also get an estimate of the quantile density for model errors in the regression model. The finite-sample density and other characteristics of averaged regression quantiles can be successfully approximated, using the saddle-point technique.

The invariance principles and strong embedding theorems, valid for empirical quantile process, naturally extend to the averaged regression quantile process. This provides a further useful tool for the statistical inference in the linear model.

Karel Kadlec

Convergence of the average cost in the case of the jump diffusions

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

Karel.Kadlec.9@seznam.cz

In this contribution, the α-stable Lévy processes are considered. The ergodic LQ stochastic optimal control problem is formulated and sufficient conditions for the convergence of the average value of the cost functional are stated and the limit of the average value of the cost functional under this sufficient conditions is computed. The optimal control is given in the feedback form and the stochastic Riccati equation corresponding to this LQ problem is derived.

Literature

[1] Applebaum D. Lévy Processes and Stochastic calcullus, Cambridge University Press, Cambridge, 2004.
[2] Barbu V. Da Prato G., Hamilton-Jacobi Equations in Hilbert Spaces. *Res. Notes in Math.*, Pitman, Boston-London, 86, 1983.
[3] Bertoin J. Lévy Processes, Cambridge University Press, Cambridge, 1996.
[4] Kallemberg O. Some time change representations of stable integrals, via predictable transformations of local martingales. *Stochastic Processes and their Applications*, North-Holland, 40, 199–223 1992.
[5] Øksendal B. Applied Stochastic Control of Jump Diffusions, Springer, New York, 2009.
[6] Peszat S., Zabczyk J. Stochastic Partial Differential Equations Driven by Lévy Processes, Cambridge University Press, Cambridge, 2006.
[7] Tang H., Wu Z. Stochastic differential equations and stochastic linear quadratic optimal control problem with Lévy process. *Jrl Syst Sci & Complexity*, 22, 122–136, 2009.

Alžběta Kalivodová[a,b], Karel Hron[a,b], Peter Filzmoser[c], Lukáš Najdekr[d]

Metoda dílčích nejmenších čtverců pro kompoziční data s aplikací v metabolomice

[a] Univerzita Palackého v Olomouci, Katedra matematické analýzy a aplikací matematiky, 17. listopadu 12, 771 46, Olomouc; [b] Univerzita Palackého v Olomouci, Katedra geoinformatiky, 17. listopadu 50, 771 46, Olomouc; [c] Department of Statistics and Probability Theory, Vienna University of Technology, Wiedner Hauptstraße 8-10, 1040 Vienna, Austria; [d] Laboratoř metabolomiky, Ústav molekulární a translační medicíny Lékařské fakulty Univerzity Palackého v Olomouci, Hněvotínská 5,779 00 Olomouc

Kalivodovaa@gmail.com

Soubor organických sloučenin, jejichž velikost je na úrovni molekul, které jsou obsaženy v daném biologickém materiálu, se nazývá metabolom. Jsou zde zahrnuty všechny organické látky přirozeně se vyskytující v metabolismu sledovaného živého organismu. Analýza metabolomu za daných podmínek se nazývá metabolomika [1]. Při kvantifikaci informací z metabolomiky mají výsledky často podobu dat nesoucích pouze relativní informaci. Vektor těchto dat má kladné složky, relevantní informace je obsažena v podílech mezi nimi, případným přeškálováním se tedy tato informace nemění. Uvedená pozorování označujeme jako kompoziční data [2], jejich statistická analýza by měla uvedené vlastnosti zohledňovat.

Metoda dílčích nejmenších čtverců je třídou metod, která se užívá k modelování vztahů mezi pozorováními (vysvětlujícími proměnnými) a skupinou tzv. latentních proměnných [3]. Můžeme ji také chápat jako kombinaci metody hlavních komponent a mnohorozměrné regrese. Tato metoda se užívá v chemometrii k mnohorozměrné kalibraci [4]. Cílem je predikovat množinu závislých proměnných na základě informace ze skupiny latentních proměnných (prediktorů). Predikce je prováděna extrakcí ortogonálních faktorů z prediktorů, cílem je maximalizovat kovarianci mezi těmito skupinami proměnných [3]. Metoda dílčích nejmenších čtverců pro kompoziční data vyžaduje specifický přístup k odhadování regresních parametrů volbou interpretovatelých ortonormálních souřadnic [5], zohledňujících jednotlivé latetní proměnné. Následně je postupně aplikována standardní regrese metodou dílčích nejmenších čtverců.

Teoretické aspekty použití kombinace metody dílčích nejmenších čtverců a logratio metodiky kompozičních dat jsou doplněny o praktický příklad z metabolomiky. Data odrážejí metabolický profil a změny metabolitů v modelovém orgamismu. Konkrétně se jedná o vzorky suchých krevních skvrn z novorozeneckého screeningu - vzorky od pacientů trpících deficitem dehydrogenáz a acyl-CoA o středně dlouhých řetězcích jsou porovnávány s kontrolními vzorky. Na tato data byla aplikována necílená metabolická analýza, ze které vzešlo přes pět set potencionálních metabolitů. Dále je provedena statistická analýza signifikance těchto metabolitů. Jsou zde porovnány aplikace standardního postupu a postupu beroucí v potaz přirozené vlastnosti kompozičních dat. Na základě podrobné analýzy dosažených výsledků lze konstatovat, že kompoziční přístup lépe identifikuje markery (charakteristické metabolity) dané nemoci.

Literatura

[1] Roux, A., Lison, D., Junot, Ch. and Heilier, J. (2011). Applications of liquid chromatography coupled to mass spectrometry-based metabolomics in clinical chemistry and toxicology: A review *Clinical Biochemistry* 44 , 119–135.
[2] Aitchison, J. (1986). *The Statistical Analysis of Compositional Data*. Monographs on Statistics and Applied Probability. Chapman & Hall Ltd., London (UK). (Reprinted in 2003 with additional material by The Blackburn Press). 416 p.
[3] Rosipal, R. and Kramer, N. (2006). *Overview and Recent Advances in Partial Least Squares*. SLSFS, Springer, 34–51.

[4] Varmuza, K. and Filzmoser, P. (2009). *Introduction to Multivariate Statistical Analysis in Chemometrics.* Taylor & Francis, New York. 336 p.
[5] J. J. Egozcue, V. Pawlowsky-Glahn (2005). *Groups of Parts and Their Balances in Compositional Data Analysis.* Mathematical Geology, 37 (7), 795–828.

Práce byla podpořena Operačním programem vzdělávání pro konkurenceschopnost - Evropský sociální fond (projekt CZ.1.07/2.3.00/20.0170 Ministerstva školství mládeže a tělovýchovy České republiky) a grantem PřF UPOL (PrF_2013_013 - Matematické modely).

Nikola Kaspříková

Některé potíže s klasifikačními modely v praxi

VŠE FIS, nám. W. Churchilla 4, 130 67 Praha 3
school@tulipany.cz

V současnosti již díky výsledkům ve statistice a strojovém učení existuje mnoho metod pro klasifikaci (jak pro režim učení s učitelem, tak pro režim učení bez učitele) a rozvoj algoritmů dál pokračuje. Je zvykem algoritmy vyvíjet a hodnotit v umělých, laboratorních podmínkách. Při řešení úloh v praxi (a zdaleka nejen ve zvláště citlivých oblastech jako třeba odhalování podvodů) ale bývají podmínky podstatně odlišné a při aplikaci algoritmů se objevují vážné problémy. Mezi takové může patřit například nepřesná představa o cenách chybných rozhodnutí nebo nedostupnost jasného návodu, jak v dané situaci zvolit optimalizační kritérium, případně podle čeho nejlépe posuzovat podobnost případů. Pokusíme se o diskuzi některých takových problémů a zkušeností s analýzami v praxi, což by mohlo napovědět něco o tom, s jakou pozorností se věnovat výkonnosti algoritmů zjišťované v umělých podmínkách.

Stanislav Katina

Analýza tvaru a obrazu: mnohorozmerné splajny a ich použitie v automatickej identifikácii, a štatistickej analýze anatomických kriviek a plôch

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

Troj-dimenzionálny biologický objekt zachytíme napr. pomocou stereo-fotogrametrických kamier alebo laserového skeneru. Tvar tohto objektu je možné popísať prostredníctvom niekoľkých miliónov triangulovaných bodov. Z nich len malá časť dostatočne dobre popisuje tvar. Nazývajú sa (semi)landmarky, ktoré môžeme rozdeliť na homologické význačné body, anatomické krivky a plochy. Tieto je nutné najprv identifikovať na všetkých objektoch náhodného výberu, aby sme ich mohli následne použiť v nejakom mnohorozmernom štatistickom modeli charakterizujúcom variabilitu tvaru alebo reprezentujúcom nejaký kauzálny vzťah. V prednáške budú predstavené niektoré metódy automatickej identifikácie kriviek a ich implementácie v jazyku R, ako napríklad diferenciálnogeometrické charakteristiky plochy, detekcia zmeny sklonu na lokálnych hlavných krivkách a vyhladzovacie splajny a ich modifikácie. Tieto metódy budú aplikované okrem iného aj na 3D obrazoch ľudských tvárí. Výsledky budú vizualizované v R knižnici rgl.

Jan Klaschka

O Blakerově konfidenčním intervalu z jiné strany

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

Blakerův konfidenční interval pro parametr binomického rozdělení [1] je jednou z „méně konzervativních“, leč exaktních alternativ ke staršímu a známějšímu Clopper-Pearsonovu intervalu [2]. Na ROBUSTu jsem o něm mluvil už dvakrát [3, 4] a vesměs šlo o problematiku jeho numerického výpočtu. Tentokrát zaměřím pozornost jiným směrem a budu se Blakerovým intervalem zabývat z hlediska metodologického.

Vos a Hudson(ová) [5] kritizovali jistou třídu oboustranných testů a konfidenčních intervalů, která zahrnuje i Blakerův interval a odpovídající test, a ukazují na příkladech, že se chovají rozporuplně: Může se stát, že nějaká data postačují k tomu, aby určitá hypotéza byla zamítnuta, ale když se datový soubor ještě zvětší, a to tak, že by podle zdravého rozumu „indicie v neprospěch hypotézy“ měly ještě zesílit, nastane opak očekávaného – data najednou k zamítnutí hypotézy nestačí.

V přednášce si ukážeme jednoduchou a výpočetně schůdnou modifikaci Blakerova konfidenčního intervalu, jež zachovává dobré vlastnosti intervalu (interval zůstává exaktní, je nadále podmnožinou Clopper-Pearsonova intervalu a závislost mezí na hladině spolehlivosti je monotónní), ale výše zmíněné rozpory odstraní.

Literatura

[1] Blaker H. (2000). *Confidence curves and improved exact confidence intervals for discrete distributions.* Canadian J. of Statistics **28**, 783 – 798.

[2] Clopper C. J., Pearson E. S. (1934). *The use of confidence or fiducial limits illustrated in the case of the binomial.* Biometrika **26**, 404 – 413.

[3] Klaschka J. (2010). *O výpočtu Blakerova konfidenčního intervalu.* ROBUST 2010. Kniha abstraktů, 20 – 21. JČMF Praha 2010.

[4] Klaschka J. (2012). *Podruhé o výpočtu Blakerova konfidenčního intervalu: Balíček BlakerCI a jiné resty.* ROBUST 2012, sborník abstraktů, 16. MFF UK Praha 2012.

[5] Vos P. W., Hudson S. (2008). *Problems with binomial two-sided tests and the associated confidence intervals.* Aust. N. Z. J. Stat. **50**, 81–89.

Lev Klebanov

Pre-limit theorems and their applications

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

levbkl@gmail.com

Finitely many empirical observations can never justify any tail behavior, so they cannot justify the applicability of classical limit theorems in probability theory. In the talk we show that instead of relying on limit theorems, one may use so-called pre-limit theorems described in the talk. These result rely not on the tail, but on the central section (the "body") of relevant distribution. Here, instead of a limiting behavior (when the number n of iid observations tends to infinity), the pre-limit theorems provide approximations of distribution functions when n is "large" but not "too large".

Jana Klicnarová

Limitní věty pro slabě závislá náhodná pole

Ef JČU ČB, KAMI, Studentská 13 370 05 České Budějovice

klicnarova@ef.jcu.cz

Existuje mnoho výsledků na téma limitních vět pro slabě závislá náhodná pole. My se v tomto příspěvku zaměříme na problematiku martingalových aproximací pro náhodná pole a na výsledky využívající technik aproximací m-závislými posloupnostmi. Ukážeme centrální limitní větu pro náhodná pole při sumaci přes obecné množiny.

Tento příspěvek je založen na společné práci s profesorem D. Volným a dr. Wangem – článek v přípravě.

Arnošt Komárek

Regrese s korelovanými intervalově cenzorovanými daty zatíženými nepřesnou klasifikací události

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

komarek@karlin.mff.cuni.cz

Základním problémem, kterým se budeme zabývat, bude regresní analýza s odezvou danou jakou čas T do nějaké události. Jestli daná událost nastala je přitom zjišťováno pouze v předem daných časových momentech, řekněme $0 = V_0 < V_1 < \cdots < V_m$. Představíme-li si pod hodnotou veličiny T čas vypuknutí nějaké (nikoliv akutní) nemoci, momenty $V_1, \ldots, V_m$ mohou reprezentovat okamžiky návštěvy lékaře či laboratorních vyšetření, kdy daná nemoc může být diagnostikována.

Při standardním intervalovém cenzorování je potom znám interval, ve kterém sledovaná událost nastala, tj. je známo $T \in (V_j, V_{j+1}]$ pro nějaké $j = 0, \ldots, m-1$ odpovídající situaci, kdy je v čase V_{j+1} zjištěn výskyt události (diagnostika nemoci), resp. $T \in (V_m, \infty)$ odpovídající stavu, že do času V_m událost ještě nenastala. Jako data reprezentující odezvu máme potom pro každou jednotku interval, ve kterém (jistě) nastala sledovaná událost.

V rámci ROBUSTního příspěvku vše zkomplikujeme (v praxi poměrně často splněným) předpokladem, že klasifikace, zda událost do okamžiku V_j nastala či nenastala, je zatížena diagnostickou chybou. To jest událost je klasifikována diagnostickým testem s nenulovou pravěpodobností falešně pozitivního, resp. negativního výsledku. Za data reprezentující odezvu na jedné jednotce lze poté brát posloupnost m jedniček a nul (jednička na $j$tém místě, jestliže diagnostický test indikuje v čase V_j, že událost již nastala, nula naopak). S perfektním diagnostickým testem by taková poslopnost měla být monotónní (tj. první jedničce předcházejí pouze nuly a dále se

vyskytují již jenom jedničky), nikoliv však u testu s nenulovou pravděpodobností falešně pozitivních, resp. negativních výsledků. Ukážeme (Bayesovský) model umožňující nejenom regresní analýzu s takovými daty (kdy odezvou je stále původní čas T do události), ale též odhad sensitivity a specificity příslušného diagnostického testu. A abychom si vše ještě malinko zkomplikovali, budeme uvažovat též situaci, kdy na jedné experimentální jednotce zjišťujeme více časů do událostí a nelze tedy automaticky předpokládat jejich nezávislost. Vše bude ilustrováno na jednom reálném medicínském problému.

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

Predikce roční spotřeby zemního plynu po ceníkových pásmech

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

konar@cs.cas.cz

V příspěvku bude představen model pro predikci počtů odběratelů zemního plynu a jejich celkových ročních spotřeb v rámci velké skupiny zákazníků (typicky zákaznického kmene distribuční společnosti) podle ceníkových pásem. Predikce je potřeba provádět vždy zhruba v polovině kalendářního roku vždy na následující kalendářní rok. Oproti klasické úloze predikce časových řad je v tomto případě situace komplikována následujícími faktory:

1. Přiřazení ceníkového pásma zákazníkovi je dáno výší jeho poslední roční spotřeby. Zákazníci tak mezi pásmy mohou migrovat, v důsledku čehož časové řady spotřeb i počtů zákazníků v jednotlivých ceníkových pásmech nejsou vzájemně nezávislé.

2. Údaje o spotřebě (a tudíž i o ceníkovém pásmu) jednotlivých zákazníků nejsou v jednotném časovém rozlišení. Zákazníci jsou odečítáni v různých časových okamžicích. Interval mezi odečty typicky délky jednoho roku, ale může být v rozpětí od několika týdnů do 18 měsíců. Zákazníci jsou navíc průběžně odečítáni v průběhu celého roku.

3. Spotřeba je teplotně závislá, teplota na následující kalendářní rok však není známa a je prakticky nepredikovatelná.

Predikční model je proto konstruován ve dvou úrovních. V první úrovni je modelována migrace zákazníků mezi pásmy jako (nehomogenní) Markovský proces s tím, že matice pravděpodobností přechodu jsou odhadovány empiricky z dostupných dat. V další úrovni je pak predikována celková roční spotřeba v daném ceníkovém pásmu za podmínky počtu zákazníků predikovaného v první úrovni a za podmínky dlouhodobě normální teploty.

Model byl navrhován, odhadován a testován na fakturačních datech distribučních společností skupiny RWE, kde má být v dohledné době nasazen do provozu jako podpůrný nástroj pro plánování.

Antonín Koubek

Časoprostorová separovatelnost a ambitové procesy

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

antonin.koubek@seznam.cz

Ambitové procesy patří mezi neseparovatelné časoprostorové procesy. Takové procesy nejsou zatím v literatuře dobře prozkoumány z důvodu velké výpočetní složitosti. V naší práci představujeme dva ambitové modely a diskutujeme možnosti posouzení časoprostorové separovatelnosti pomocí F funkce z replikovaných pozorování procesu. Diskutujeme možnosti statistického testování hypotézy, že je daný model separovatelný. Porovnáváme naše výsledky s jedním časoprostorově separovatelným modelem popsaným ve článku Møller, Ghorbani (2012), kde bylo použití F funkce navrženo. Pro tyto modely odvozujeme teoretické hodnoty F funkce a také teoretické hodnoty párové korelační funkce. Uvádíme také výsledky simulací a numerické hodnoty zkoumaných funkcí. Tím je umožněna kontrola správnosti našich výsledků.

Pavel Kříž

Probability limit identification functions

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

pavel-kriz@post.cz

The poster presents the concept of a PLIF - a function identifying probabilty limits. Main theoretical results regarding existence and (non-)measurability are summarized and possible applications in estimation theory and stochastic analysis are outlined.

Michal Kulich

Odhadování incidence HIV z průřezových dat

KPMS MFF UK, Sokolovská 83, 186 75 Praha 8

kulich@karlin.mff.cuni.cz

Incidence HIV se standardně odhaduje sledováním kohorty zdravých jedinců a zaznamenáváním nově nakažených případů. V některých aplikacích však dlouhodobé sledování a opakované testování není proveditelné. K dispozici jsou pouze průřezová data obsahující informaci o tom, který jedinec je nakažený, a hodnoty laboratorních vyšetření, které jsou nějakým způsobem korelované s dobou od nakažení. Pomocí laboratorních měření se určí skupina HIV-pozitivních pacientů, kteří jsou považováni za „nedávno nakažené“ a z jejich počtu se odhadne incidence.

Při tomto přístupu není obecně možné dostat konsistentní odhad incidence, nicméně lze optimalizovat pravidlo determinující nedávno nakažené jedince tak, že výsledné vychýlení je pro praktické aplikace tolerovatelné. Ukážeme, jakým způsobem lze provést takovou optimalizaci na validačním souboru a jak lze vyhodnotit vlastnosti získaných odhadů. Tato metodika byla použita při provedení primární analýzy randomizované studie Projekt ACCEPT.

Petra Kynčlová[1,3], Peter Filzmoser [1], Karel Hron[2,3]

Aplikace $\mathcal{T}$-prostorů při modelování kompozičních časových řad

[1]Department of Statistics and Probability Theory, Vienna University of Technology, Austria; [2]Department of Mathematical Analysis and Applications of Mathematics, Palacký University, Olomouc, Czech Republic; [3]Department of Geoinformatics, Palacký University, Olomouc, Czech Republic

kynclova.petra@gmail.com

Mnohorozměrné časové řady, které modelují relativní podíly částí na celku, se nazývají kompoziční časové řady. V praxi se pak často jedná o proporcionální data, jejichž výběrovým prostorem je simplex. Z toho důvodu je nejdříve potřeba tyto časové řady vyjádřit (pomocí tzv. logratio transformací) v souřadnicích standardního euklidovského prostoru, kde lze následně aplikovat vybraný mnohorozměrný model. Přitom jak výsledný model na simplexu (získaný zpětnou transformací), tak odpovídající předpovědi nezávisí na volbě konkrétní logratio transformace [2].

Kompoziční data jsou obvykle reprezentována jako mnohorozměrná data popisující části daného celku, a tedy nesou výhradně relativní informaci [1]. V případě modelování kompozičních časových řad může ovšem absolutní informace o celkovém množství hrát důležitou roli, např. při predikci budoucích hodnot. Z tohoto důvodu se ukázalo jako vhodné využití teorie tzv. $\mathcal{T}$-prostorů, která definuje rozšířený prostor $\mathcal{T} = \mathbb{R}_+ \times \mathcal{S}^D$ a jenž umožňuje modelovat současně podíly mezi jednotlivými složkami a úhrnné hodnoty těchto složek [5].

Cílem příspěvku je popsat metodiku modelování kompozičních časových řad vzhledem k Aitchisonově geometrii jako nedílnou součást $\mathcal{T}$-prostorů při aplikaci vektorového autoregresního modelu. Na reálném příkladě pak bude demonstrováno, že výběrem konkrétní izometrické logratio (ilr) transformace [3] lze docílit patřičné interpretace výsledků v souřadnicích, včetně možnosti testování Grangerovy kauzality uvnitř modelu [4].

Literatura

[1] Aitchison, J. (1986). *The Statistical Analysis of Compositional Data.* Monographs on Statistics and Applied Probability. Chapman and Hall Ltd., London, UK.

[2] Barceló-Vidal, C., Aguilar, L., Martín-Fernández, J. A. (2011). Compositional VARIMA time series. In V. Pawlowsky-Glahn and A. Buccianti, eds., *Compositional Data Analysis. Theory and Applications.* John Wiley & Sons, Chichester, pp. 87–103.

[3] Egozcue, J.J., Pawlowsky-Glahn, V., Mateu-Figueraz, G. (2003). Isometric logratio transformations for compositional data analysis. *Mathematical Geosciences 35*(3), 279–300.

[4] Lütkepohl, L. (2005). *New Introduction to Multiple Time Series Analysis.* Springer, Berlin.

[5] Pawlowsky-Glahn, V., Egozcue, J.J., Lovell, D. (2013). The product space $\mathcal{T}$ (tools for compositional data with a total). In K. Hron, P. Filzmoser and M. Templ, editors, *Proceedings of CoDaWork'13, The 5th Compositional Data Analysis Workshop.* Vorau, Austria.

Tato práce vznikla za podpory Operačního programu vzdělávání pro konkurenceschopnost - Evropský sociální fond (projekt CZ.1.07/2.3.00/20.0170 Ministerstva školství mládeže a tělovýchovy České republiky).

Petr Lachout

Poznámka k zápisu náhodných posloupností pomocí polynomů

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

lachout@karlin.mff.cuni.cz

Tato poznámka se týká zápisu náhodné posloupnosti, jako řešení rovnic zadaných polynomy. Upozorňuje na to, že zápis není jednoznačný. Stejnou posloupnost lze vyjádřit různými soustavami rovnic, které jsou kvalitativně odlišné a přitom jejich shoda s daty je téměř identická. Problematika bude ukázána na příkladě ARMA procesu. Při výkladu bude využita obecná algebra nekonečných posloupností čísel indexovaných přirozenými čísly.

Matúš Maciak, Ivan Mizera

Change-point estimation and inference in nonparametric regression using different regularization concepts

University of Alberta, Edmonton, Kanada

maciak@ualberta.ca

Classical regression techniques require a smoothness assumption to be satis

ed. It makes the theoretical justi

cation easier and the model more straightforward to interpret. In many situations however, statisticians need to deal with more complex dependence structures where the underlying functional form is non-smooth, or even discontinuous. Such models are in statistics referred to as change-point models as locations where the smoothness (continuity) assumption is not satisfied are com- monly said to be change-points. Unfortunately, many existing methods require a prior knowledge for the location of change-points in a model, which can be quite limiting in practical situations.

We will propose a new approach to change-point estimation in regression: the main advantage of our method is that it introduces a fully data-driven approach with no requirement on prior knowledge for change-point locations. It combines nonpara- metric regression estimation with different concepts of an L1-norm regularization. Different alternatives are proposed, a proper statistical inference is discussed and theoretical results are derived. Finite sample performance is investigated using simulated data and real examples as well.

Ivan Mizera

Využitie skúsenosti v predikcii: Empirické bayesovské metódy, kvalitatívne ohraničenia a konvexná optimalizácia

University of Alberta, Edmonton, Kanada

imizera@yahoo.com

V prednáške bude podaný prehľad niektorých aktuálnych metód použiteľných v predikcii s využitím skúsenosti („empirical Bayes, náhodné efekty“), v jednoduchých hierarchických modeloch („mixture models“) - problém, ktorý sa v poslednej dobe tešil znovu pozornosti ako alternatíva k metódam typu „atomic pursuit/lasso“. Metódy majú neparametrický charakter, no napriek tomu nevyžadujú nastavenie špeciálnych parametrov, vďaka istým špecifikám ako napr. využitiu kvalitatívnych ohraničení („shape constraints“), a umožňujú efektívnu implementáciu použitím metód modernej konvexnej optimalizácie. Ak zostane čas, budú ilustrované na príkladoch z populárnych športov.

Stanislav Nagy

Konzistencia hĺbky funkcií II

MFF UK, KPMS, Sokolovská 83, 186 75 Praha 8; KU Leuven, Dept. of Mathematics

nagy@karlin.mff.cuni.cz

Konceptu hĺbky nekonečnerozmerných (funkcionálnych) dát bolo v poslednej dobe v literatúre venované množstvo pozornosti. Okrem známych prístupov pomocou rôznych modifikácií *hĺbok integrálneho typu* a *pásových hĺbok* nedávno Mosler a Polyakova [3] predstavili novú hĺbku *infimálneho typu*. V príspevku budeme vyšetrovať podmienky za ktorých je výberová verzia takýchto funkcionálov konzistentným odhadom svojho populačného náprotivku. To je samozrejme nutnou podmienkou ich zmysluplného využitia v aplikáciách.

Pokračujúc smerom naznačeným v príspevku „Konzistencia hĺbky funkcií“ (ROBUST 2012) ukážeme, že okrem pásových hĺbok ani hĺbky infimálneho typu nie sú univerzálne konzistentné (t.j. nie sú konzistentné pre každú pravdepodobnosť). Ďalej ukážeme, že ani pôvodný dôkaz silnej konzistencie hĺbok integrálneho typu Fraimana a Munizovej [2, Theorem 3.1] nefunguje.

Pokúsime sa viesť dôkaz iným smerom a tiež nájsť podmienky, za ktorých by sme mohli zaručiť konzistenciu infimálnych hĺbok. V prvom prípade opravujeme pôvodné tvrdenie [2, Theorem 3.1], odstraňujeme jeho zbytočné predpoklady a zároveň zosilňujeme výsledok o konzistencii iných hĺbok integrálneho typu Cuevasa a Fraimana [1, Theorem 2]. V prípade hĺbok infimálneho typu zatiaľ, pokiaľ nám je známe, žiadny podobný výsledok v literatúre neexistuje.

Literatúra

[1] Cuevas, A. and Fraiman, R.: 2009, ‘On depth measures and dual statistics. A methodology for dealing with general data’. *J. Multivariate Anal.* **100**(4), pp. 753–766.

[2] Fraiman, R. and Muniz, G.: 2001, ‘Trimmed means for functional data’. *Test* **10**(2), pp. 419–440.

[3] Mosler, K. and Polyakova, Y.: 2013, ‘General notions of depth for functional data’. *arXiv preprint arXiv:1208.1981*

Radim Navrátil, Jana Jurečková

Pořadové testy v regresi při rušivé heteroskedasticitě

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

navratil@karlin.mff.cuni.cz, jurecko@karlin.mff.cuni.cz

Homoskedasticita je často mlčky předpokládána při analýze lineárního modelu, ať už klasickými či robustními metodami. Abychom se vyhnuli negativním následkům při ignorování heteroskedasticity, můžeme buďto testovat její přítomnost ještě předtím, než začneme vlastní statistickou inferenci, nebo se snažíme najít přístup, který je invariantní vůči heteroskedasticitě.

V příspěvku navrhneme v lineárním regresním modelu s heteroskedastickými chybami neparametrické testy o regresi za přítomnosti rušivé heteroskedasticity a testy heteroskedasticity za přítomnosti rušivé regrese. Oba typy testů jsou založeny na vhodných ancilárních statistikách, není tedy nutné odhadovat rušivé parametry modelu, na rozdíl od testů navrhovaných v literatuře. Simulační studie stejně jako aplikace na reálná data ukazuje dobré chování navržených testů.

Petr Novák

Odhady základního rizika v regresních modelech oprav

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

novakp@karlin.mff.cuni.cz

Pozorujeme nezávislá zařízení podléhající opotřebení a pomocí vhodných regresních modelů se snažíme popsat vliv jejich průběžných oprav a údržby na rozdělení doby do selhání. Nejčastěji používané modely, jako je Coxův model proporcionálního rizika nebo model zrychleného času, popisují vliv regresorů na určitou základní rizikovou funkci. Tu je potřeba buď vhodně parametrizovat, nebo odhadnout neparametricky. V příspěvku se zaměřujeme na metody porovnávání a testování hypotéz o tvaru základního rizika v modelech oprav a předvádíme jejich využití.

Zbyněk Pawlas

Statistika Poissonových modelů pro sjednocení kruhů

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

pawlas@karlin.mff.cuni.cz

V řadě oborů se vyskytují geometrické struktury, které je moľné modelovat pomocí sjednocení náhodně rozmístěných kruhů v rovině (obecněji koulí v d-rozměrném euklidovském prostoru). Základním modelem je Booleův model, ve kterém se rozmístění kruhů řídí Poissonovým procesem. Uvaľujeme stacionární Booleův model kruhů, který je definován jako

$$Z := \bigcup_{n\geq 1} B(\xi_n, R_n),$$

kde $B(x,r)$ je kruh se středem x a poloměrem r, $\Phi := \{\xi_n, n \geq 1\}$ je stacionární Poissonův bodový proces v rovině a $\{R_n, n \geq 1\}$ je posloupnost nezávislých stejně rozdělených náhodných veličin, které jsou nezávislé na Φ. Takovýto model je určen intenzitou γ bodového procesu Φ středů kruhů a pravděpodobnostním rozdělením $\mathbb{G}$ poloměrů kruhů. Důleľitým statistickým problémem je, jak získat informaci o γ a $\mathbb{G}$ z pozorování jedné realizace náhodné mnoľiny Z v omezeném okně W. V přednášce zmíníme existující postupy pro odhad γ i $\mathbb{G}$ a představíme novou třídu neparametrických odhadů rozdělení $\mathbb{G}$.

Michal Pešta

Trojuholníkové dáta, podmienené najmenšie štvorce, pseudovierohodnosť a kopule

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

Michal.Pesta@mff.cuni.cz

Trojuholníkové dáta pozostávajú z n kópií stochastického procesu, kde prvá realizácia má práve n pozorovaní a každá ďalšia o jedno menej, až posledná má len jedno pozorovanie. Predstavíme triedu modelov pre podmienenú strednú hodnotu a rozptyl stochastického procesu, kde závislostná štruktúra medzi jednotlivými pozorovaniami procesu bude reprezentovaná kopulou.

Podmienené najmenšie štvorce v kombinácii s pseudovierohodnosťou použijeme na odhad parametrov tried modelov pre trojuholníkové dáta. Využitím mixingalových vlastností ukážeme konzistenciu odhadov parametrov.

Motivácia pre stochastické modely na trojuholníkové dáta pochádza z neživotného poistenia, takže záverom aplikujeme teoretické prístupy na reálne dáta.

Literatúra

[1] Hudecová Š. a Pešta M. (2013) Modeling dependencies in claims reserving with GEE. *Insurance: Mathematics and Economics*, 53, 786 – 794.

[2] Pešta, M., Okhrin, O. (2013): Conditional least squares and copulae in claims reserving for a single line of business. *Submitted to Insurance: Mathematics and Economics*, http://arxiv.org/abs/1306.4529.

Samuel Rosa, Radoslav Harman

Optimal trend resistant experimental designs for comparing treatments with a control

Comenius University in Bratislava, FMPI, Mlynská dolina 842 48 Bratislava 4

srrs.sk@gmail.com

Suppose that we intend to perform a sequence of independent trials, each with one of v treatments. Let the first treatment be a control and let the effects of the treatments be denoted by $\tau_1, \ldots, \tau_v$. The mean value of the response of each trial is assumed to be equal to the sum of the effect of the treatment selected for the trial, and the effect of a nuisance time trend. We give a class of optimal approximate designs for the estimation of a set of contrasts $\tau_2 - \tau_1$, ..., $\tau_v - \tau_1$, with respect to any Kiefer's ϕ_p-optimality criterion, $p \in [-\infty, 0]$, including the criteria of D-, A- and E-optimality. The results can be used to construct a lower bound on efficiency of any exact trend design.

Zuzana Rošťáková

Stochastické modelovanie vel'kých škôd v poisťovníctve

Ústav merania SAV, Bratislava

zuzana.rostakova@gmail.com

Veľké škody tvoria v neživotnom poistení len malú časť z celkového počtu poistných udalostí. Na druhej strane, ich príspevok v konečnej sume poistných plnení je pomerne vysoký. Vo všeobecnosti ich môžeme chápať ako extrémne pozorovania. Cieľom tohto príspevku bolo vytvorenie modelu, na základe ktorého by sa dala odhadnúť výška budúcich škôd.

Na úvod sme uviedli niektoré grafické metódy, ktoré je možné využiť pri zisťovaní, či dáta pochádzajú z rozdelenia s ťažkým chvostom, alebo nie. Vhodným nástrojom na modelovanie ťažkých chvostov je zovšeobecnené rozdelenie veľkých hodnôt (GEV distribution). Toto rozdelenie umožňuje definovať „prah“ ako hranicu medzi veľkými a zanedbateľnými škodami.

Pri tvorbe modelu sme sa zamerali predovšetkým na metódu POT – Peaks Over Threshold, ktorá je založená práve na voľbe vhodného prahu pomocou zovšeobecneného Paretovho rozdelenia. Na odhady potrebných parametrov sme použili metódu maximálnej vierohodnosti a váženú momentová metódu.

Pomocou vybudovanej teórie sme na záver vytvorili model pre dve sady reálnych pozorovaní a otestovali sme jeho schopnosť odhadnúť výšku budúcich škôd.

Vladimíra Sečkárová[1,2], Radka Sabolová[1]

I-divergence based statistical inference in exponential family

[1]MFF UK, KPMS, Sokolovská 83, 186 75 Praha 8
[2] ÚTIA AV ČR, Pod Vodárenskou věží 4, Praha 8

seckarov@karlin.mff.cuni.cz, sabolova@karlin.mff.cuni.cz

The I-divergence [2] represents a tool for statistical inference about an unknown parameter γ of a probability distribution satisfying the following conditions: (i) it belongs to the regular exponential family and (ii) possesses the covering property [1]. We exploit the use of the I-divergence from two different views. Firstly, we propose a graphical method for I-divergence based testing of parameter γ exploiting the cumulative distribution function, quantiles of the I-divergence and quantiles of the uniform distribution. The description is followed by the application to simulated exponentially distributed data. Secondly, we discuss the decomposition of the I-divergence into two independent elements, both having statistical interpretation in hypothesis testing. The aim of this part is to show the decompositions for several members of the exponential family, namely the exponential, gamma and Pareto distribution.

Literatura

[1] Pázman, A. (1993): *Nonlinear statistical Models.* Kluwer Acad. Publ., Dordrecht, chapters 9.1 and 9.2.
[2] Stehlík, M. (2003) Distributions of exact tests in the exponential family. *Metrika* **57** 145 – 164.

The authors have been partially supported by the Aktion grant Austria-Czech Republic and by the grant SVV-2013 - 267 315.

Marina Stecenková

Klasifikace vzorů v EEG signálu

VŠE FIS, nám. W. Churchilla 4, 130 67 Praha 3

marina.stecenkova@vse.cz

Tento příspěvek se zabývá klasifikací pro projekt „rozhraní mozek-počítač“ (BCI, Brain-Computer Interface), který je založen na čtení a analyzování EEG signálů. Podstata BCI projektů spočívá ve vytvoření komplexního systému pro přímou interakci mezi mozkem a vnějším technickým zařízením. Hlavním prvkem takového BCI je klasifikátor vzorů v EEG signálu, které odpovídají provádění různých duševních úkolů. EEG je vícerozměrný, velmi nestabilní signál obsahující mnoho šumu, proto je klasifikace těchto vzorů obtížný úkol. Cílem tohoto příspěvku je představit Bayesovský klasifikátor, který bude validován na reálných fyziologických experimentech. Problém dimenzionality a vysokého obsahu šumu v datech bude vyřešen pomocí analýzy nezávislých komponent (ICA).

David Stibůrek

Testování hypotéz parametru driftu u stochastických procesů

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

david.stiburek@gmail.com

Pro statistické vyšetřování parametru driftu a s rušivým parametrem σ u Wienerova procesu W_t s deterministickým driftem $Y_t = ad(t) + \sigma W_t$ (pro $d(t)$ známé) existuje řada možností. Toto vyšetřování můžeme například založit na vlastnostech lineárním modelu aplikovaného na přírůstky mezi jednotlivými pozorováními daného procesu. U testování hypotéz parametru driftu a je však časově efektivnější použít inverzní metody založené na čase prvního výstupu z předem stanoveného intervalu a do konečně zvoleného času. Tyto metody mohou být přímo použity pro Wienerův proces s konstantním driftem $Y_t = at + W_t$. Díky univerzálnosti tohoto postupu je užití takovýchto metod široké. Příklady dalších procesů s užitím a srovnáním průměrných časů rozhodování je graficky ilustrováno.

Zdeněk Šulc

Porovnání nových přístupů v oblasti měr podobnosti pro kategoriální data

VŠE FIS, nám. W. Churchilla 4, 130 67 Praha 3

zdenek.sulc@vse.cz

Příspěvek hodnotí vybrané míry podobnosti, které byly v nedávné době navrženy pro účely shlukování objektů charakterizovaných nominálními proměnnými. Tyto míry byly ověřovány na několika datových souborech. Získané shluky jsou hodnoceny podle různých indexů zahrnujících normalizovaný Giniho koeficient a normalizovanou entropii, modifikované pseudo F indexy založené na Giniho koeficientu a entropii. Výsledná hodnocení jsou porovnávána s výsledky dosaženými s využitím běžně používaných měr, konkrétně s koeficientem prosté shody a věrohodnostní mírou vzdálenosti. Provedené experimenty naznačují, že některé z nedávno navržených měr podobnosti poskytují výrazně lepší shluky v porovnání s běžnými mírami, zejména s koeficientem prosté shody.

Petr Veverka

On near-optimal necessary and sufficient conditions for forward-backward stochastic systems with jumps

FJFI ČVUT, Trojanova 12, 120 00 Praha 2

petr.veverka@fjfi.cvut.cz

In the talk, the necessary and sufficient conditions for near-optimality of controlled nonlinear stochastic systems will be given. In our case, the controlled state process considered is governed by Forward-Backward Stochastic Differential Equation with jumps. The result is a joint work of P.V., M. Hafayed and S. Abbas.

Ondřej Vencálek

Využití hloubky dat pro klasifikaci – globální a lokální přístupy

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

ondrej.vencalek@upol.cz

Uspořádání bodů ve vícerozměrném prostoru $\mathbb{R}^d$ vzhledem k nějakému rozdělení pomocí hloubky dat může být základem pro řešení mnoha statistických úloh. Jednou z nich je i úloha klasifikace – tvorby pravidla pro zařazení nového pozorování do jedné z $K \geq 2$ skupin, jejichž reprezentanty pozorujeme. Během posledních přibližně deseti let bylo navrženo mnoho klasifikátorů využívajících hloubku dat. Cílem příspěvku je přehledně shrnout práci v oblasti klasifikace na základě hloubky dat a ukázat nové trendy v této oblasti.

Typický postup nalezení klasifikátoru využívajícího hloubku dat je dvoukrokový. Prvním krokem je výpočet hloubek původních pozorování vůči jednotlivým skupinám bodů tréningové množiny. Jde tedy o zobrazení $\mathbb{R}^d \to [0,1]^K$. Jelikož typicky (i když ne nutně) platí, že $K < d$, můžeme tento krok chápat jako redukci dimenze úlohy. Navíc budeme nadále pracovat s kompaktní množinou $[0,1]^K$. Druhým krokem je nalezení vhodného klasifikátoru na prostoru $[0,1]^K$.

Různé klasifikátory se liší v tom, jak realizují dva výše uvedené kroky. Rozdílnost v prvním kroku plyne z různosti hloubek v tomto kroku použitých. Může být použita např. poloprostorová, projekční, zonoidová či L_1 hloubka. Kterákoliv z výše uvedených hloubek určitého bodu je globální charakteristikou tohoto bodu udávající míru jeho centrality vůči nějakému rozdělení či skupině bodů. Je však možno také použít některou z lokálních hloubek. Rovněž druhý krok může být realizován různě. Některé procedury se dají označit jako globální, jiné jako lokální. Za globální označíme ty procedury, kde k zařazení nového pozorování porovnáváme jeho hloubky s hloubkami všech bodů tréningové množiny. Naopak lokální procedury jsou typicky založené na porovnávání s blízkými sousedy (z hlediska hloubek).

Globálnost a lokálnost postupů v obou krocích s sebou přináší jisté výhody i nevýhody. Globální postupy mohou těžit z globálních vlastností uvažovaných rozdělení, jako je unimodalita či symetrie. Lokální postupy se zdají být vhodnou alternativou tam, kde s výše uvedenými globálními vlastnostmi nelze počítat. Lépe tedy vyhovují představě neparametrického postupu jakožto postupu s minimálním množstvím předpokladů o tvaru uvažovaných rozdělení.

Jan Ámos Víšek

Diagnostics of the robustified least squares

FSV UK, Smetanovo nábřeží 6, 110 01 Praha 1

visek@fsv.cuni.cz

Let's consider the regression model as

$$Y_i = X_i'\beta^0 + e_i, \qquad i = 1, 2, ..., n.$$

Put for any $\beta \in R^p$ $r_i(\beta) = Y_i - X_i'\beta$ and $r^2_{(1)}(\beta) \le r^2_{(2)}(\beta) \le ... \le r^2_{(n)}(\beta)$. Then the *least weighted squares* $\hat\beta^{(LWS,n,w)}$ (*LWS*) is given as ([7],[9]) Special cases are the *ordinary least squares*, the *least median of squares* and the *least trimmed squares*.

$$\hat\beta^{(LWS,n,w)} = \arg\min_{\beta\in R^p} \sum_{\ell=1}^{n} w_\ell\, r^2_{(\ell)}(\beta).$$

Let $w_\ell = w\big(\frac{\ell-1}{n}\big)$, $w : [0,1] \to [0,1]$, continuous, nonincreasing and $F^{(n)}_\beta(v)$ be the e. d. f. of $|r_i(\beta)|$. Straightforward technicalities show that $\hat\beta^{(LWS,n,w)}$ is one of solutions of the *normal equations*

$$\sum_{i=1}^{n} w(F^{(n)}_\beta(|r_i(\beta)|))\, X_i\, (Y_i - X_i'\beta) = 0$$

where $F^{(n)}_\beta(v)$ is the empirical d. f. of the absolute values of residuals. It immediately inspires the proposals of the *instrumental weighted variables (IWV)* and the *weighted total least variables (WTLS)* - robustifications of the classical *IV* and *TLS*, respectively ([3] or [8]). Then $\sup_{v\in R^+} \sup_{\beta\in R^p} \sqrt{n}\left|F^{(n)}_\beta(v) - F_\beta(v)\right| = \mathcal{O}_p(1)$ (with $F_\beta(v) = P(|Y_1 - X_1'\beta| < v)$, see [10]) allows to study the theoretical properties of these estimators. We have addressed many topics concerning *LWS, IWV* and *WTLS* ([1] - [6]) but the basic problems of significance of explanatory variables or testing submodels were not yet considered, in the sense of [9] or [11]. That is the topic of the present contribution.

Literature

[1] Čížek P. (2008) General trimmed estimation: Robust approach to nonlinear and limited dependent variable models. *Econometric Theory 24(6), 1500-1529.*
[2] Čížek P. (2010). Semiparametrically weighted robust estimation of regression models. *CSDA 55(1), 774-788.*
[3] Franc J. (2010) Robustifying the total least squares. *preprint.*
[4] Kalina J. (2004) Durbin-Watson test for least weighted squares. *COMPSTAT 2004, 1287 - 1294.*
[5] Mašíček L. (2004) Optimality of the least weighted squares estimator. *Kybernetika 40, 715-734.*
[6] Skuhrovec J. (2010) Bootstraping least weighted squares. *MME, 554 - 559.*
[7] Víšek J.Á. (2000) Regression with high breakdown point. *ROBUST 2000, ed. J. Antoch, 324 - 356.*
[8] Víšek J.Á. (2009) Consistency of the instrumental weighted variables. *AISM 61, 543 - 578.*
[9] Víšek J.Á. (2011) Consistency of the least weighted squares under heteroscedasticity. *Kybernetika 47 , 179-206.*
[10] Víšek J.Á. (2011) Empirical distribution function under heteroscedasticity. *Statistics 45, 497-508.*
[11] Welsh A.H., Ronchetti E. (2002) A journey in single steps: robust one-step M-estimation in linear regression. *JSPI 103, 287 - 310.*

The research was supported by the Czech Science Foundation project No. 13-01930S

Petr Volf

On competing risks and problem of identification

Institute of Information Theory and Automation, Czech Academy of Sciences, Pod Vodárenskou věží 2, 182 08 Praha 8

volf@utia.cas.cz

The contribution deals with the problem of competing risks (of competing events) in the statistical survival analysis. In such a setting, only the first event occurs, the others are 'censored'. Further, the potential occurrence of competing events may be dependent. It is known that, in general, the right model is not identifiable from observed data. We recall the notion of incidence function and the methods of statistical incidence analysis. Then we study the relationship between marginal, joint and incidence distributions of events. We consider a joint model based on a copula. With the aid of an example we show that the assumption about the form of copula enables us to estimate underlying distribution of competing events.

The situation can be better when our information is richer thanks to dependence of data on observed covariates. We shall recall certain results concerning the Cox's and AFT regression model. As the proof given in literature [1] is rather 'intuitive' than exact, we shall demonstrate identifiability of model with the aid of artificial examples.

Literature

[1] Heckman J.J., Honoré B.E.: The identifiability of the competing risks model. Biometrika **76**, 325 – 330, 1989.

Gejza Wimmer[1], Viktor Witkovský[2]

A family of transformed Lambert $W \times Gamma$ random variables with applications

[1]Matematický ústav SAV, Bratislava; [2]Ústav merania SAV, Bratislava

wimmer@mat.savba.sk, witkovsky@savba.sk

The Lambert $W \times F$ random variables and families of their distributions have been suggested and studied by G.M. Georg in [1], as an useful tool for modeling and analyzing skewed and heavy tailed distributions. Here we shall focus on related, however different class of random variables (suitable transformations of Lambert $W \times F$ random variables), which naturally appears in statistical likelihood based inference of normal random variables:

1. in deriving the exact likelihood ratio statistic for the variance of a normal distribution;
2. in testing the simple null hypothesis on all parameters of the linear regression model with normally distributed errors;
3. in searching for exact simultaneous confidence regions for variance components.

Here we present the general shape of this distributions, its basic properties and also the special case of this distributions obtained in three above mentioned basic statistical procedures.

Literature

[1] Georg G.M. Lambert W random variables - a new family of generalized skewed distributions with applications to risk estimation. *The Annals of Applied Statistics* **5**(3), (2011), 2197 – 2230.

The work was supported by the Slovak Research and Development Agency, grant APVV-0096-10 and by the Scientific Grant Agency of the Ministry of Education of the Slovak Republic and the Slovak Academy of Sciences, grant VEGA 2/0038/12.

Viktor Witkovský

Poznámky o niektorých výpočtových aspektoch tradičných testov o pevných a náhodných efektoch v lineárnych zmiešaných modeloch

Ústav merania SAV, Bratislava

witkovsky@savba.sk

Metódy štatistickej inferecie pre testovanie hypotéz a konštrukciu konfidenčných oblastí pre pevné a náhodné efekty v zmiešaných lineárnych modeloch sú obyčajne zaožené na (približných) testoch pomerom vierohodností, alebo na (približných) testoch založených na testovacích štatistikách Waldovho typu. Napriek tomu, že tieto metódy sú dobre známe a implementované v algoritmoch štatistických počítačových balíkov určených na analýzu dát pomocou takýchto modelov, potreba analyzovať rozsiahle dáta prináša technické problémy ako takéto metódy efektívne implementovať (napr. `nlme`, `lme4` v `R`, `Proc MIXED`, `GLIMMIX` a `HPMIXED` v prostredí `SAS`, resp. `fitlme` v prostredí `MATLAB`).

V tomto príspevku stručne popíšeme štandardne používané metódy testovania hypotéz pomocou testovacích štatistík Waldovho typu a metódy aproximácie ich rozdelenia za platnosti nulovej hypotézy (aproximácia pomocou metódy *Satterthwaite-Fai-Cornelius* a *Kenward-Roger*) s dôrazom na niektoré výpočtové aspekty implementovania týchto metód.

Literatúra

[1] Witkovský V. Estimation, testing, and prediction regions of the fixed and random effects by solving the Henderson's mixed model equations. Measurement Science Review 12 (6), 2012, 234 – 248.

Práca vznikla vďaka podpore grantov APVV-0096-10, SK-AT-0025-12, VEGA 2/0038/12 a VEGA 2/0043/13.

Markéta Zikmundová

Bodové procesy úseček

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

zikmundm@karlin.mff.cuni.cz

V příspěvku se zabýváme bodovými procesy úseček, jejich rozšířením o vzájemné interakce a simulací takových modelů. Jsou zkoumány odhady parametrů modelu a číselných charakteristik procesu úseček.

Robert Zůvala, Eva Fišerová

Modelování sesuvu svahu v Halenkovicích pomocí metody kriging

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

r.zuvala@seznam.cz

Kriging patří mezi geostatistické interpolační metody, které slouží k modelování prostorových dat. Od doby, kdy byla tato metoda poprvé popsána (Matheron, 1963), se těší značnému zájmu a prodělala dlouhý vývoj. Termín kriging v sobě dnes zahrnuje celou řadu interpolačních technik, jejichž cílem je snaha o optimální lineární predikci. Základní technikou, kterou se zde budeme zabývat, je obyčejný kriging. Obyčejný kriging vychází z předpokladu stacionarity náhodného procesu, kdy korelace mezi dvěma náhodnými veličinami závisí pouze na jejich prostorové vzdálenosti. Variabilitu dat lze popsat pomocí semi-variogramu. Mezi nejběžnější modely semi-variogramu patří lineární, sférický, exponenciální nebo gaussovský. Princip modelování pomocí obyčejného krigingu je založen na váženém průměru sousedních hodnot, přičemž váhy jsou optimalizovány pomocí semi-variogramu. Obyčejný kriging, stejně jako další typy krigingu, poskytuje nejlepší nestranné lineární odhady.

Cílem příspěvku je demonstrovat základní principy krigingu při modelování sesuvu svahu poblíž obce Halenkovice ve Zlínském kraji.

Literatura

[1] Amstrong M. Basic linear geostatistic. Springer, heidelberg, 1998.

[2] Cressie N.A.. Statistics for spatial data. J. Wiley & Sons, New York, 1993.

[3] Matheron G. Pinciples of Geostatistics, Economic Geology, 1963.

Marta Žambochová

Dvoufázový způsob vytváření nekonvexních shluků využívající metody k-průměrů

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

marta.zambochova@ujep.cz

Jedním ze známých nedostatků metody k-průměrů je neschopnost hledání nekonvexních shluků. V příspěvku je popsána a analyzována možnost, jak tento nedostatek odstranit, nebo alespoň zmírnit a jak tato nová vlastnost algoritmu negativně ovlivňuje pozitiva původní výhody metody k-průměrů.