ROBUST 2018

Sborník abstraktů

Aneta Andrášiková[1], Eva Fišerová[1], Silvie Bělašková[2]

Chování silofunkcí testů v Coxově modelu proporcionálních rizik

[1]Univerzita Palackého v Olomouci, PřF, KMAAM, 17. listopadu 1192/12, 771 46 Olomouc
[2]Fakultní nemocnice u sv. Anny v Brně, Mezinárodní centrum klinického výzkumu, Pekařská 53, 656 91 Brno

andrasikova.aneta@gmail.com, eva.fiserova@upol.cz, silvie.belaskova@fnusa.cz

S analýzou přežívání se můžeme setkat v širokém spektru odvětví. Atraktivní využití nachází zejména v medicíně, kde umožňuje posouzení účinnosti daných léčebných postupů. Vztah mezi dobou přežívání (časem nastoupení sledované události) a zvolenými prediktory lze popsat pomocí Coxova modelu proporcionálních rizik. Statistická významnost prediktorů se ověřuje pomocí testu poměrem věrohodnosti, Waldova testu a skórového testu. Všechny tři uvedené testy jsou asymptotické, a proto jsou pouze přibližné. Cílem příspěvku je studium chování těchto testů pro malé datové soubory. V rozsáhlé simulační studii budeme sledovat chování silofunkcí uvedených testů v závislosti na poměru censorovaných pozorování, na rozsahu datového souboru a typu rozdělení pravděpodobnosti funkce základního rizika.

Katarína Bartošová

Klasifikácia zašumených dát pre rastúci počet vstupných premenných

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

katarina.bartosova@savba.sk

Robustnú klasifikačnú metódu, pri ktorej budeme predpokladať elipsoidálny model šumu vstupných dát, aplikujeme na analýzu dychu. Budeme sa venovať výberu vstupných premenných, teda výberu prchavých organických zložiek vydychovaných plynov, na základe Youdenovho indexu. Vyhodnotíme vplyv rastúceho počtu vstupných premenných na výsledky klasifikácie.

Poďakovanie: Táto práca bola podporovaná grantom APVV-15-0295 a VEGA 2-0054-18.

Jaromír Běláček

Odhady budoucích počtů pojištěnců VZP na bázi jednoletých časových řad

Všeobecná zdravotní pojišťovna, Oddělení strategických analýz, Orlická 4, 130 00 Praha 3

jaromir.belacek@vzp.cz

Východiska a cíle: Predikovat budoucí početní stavy pojištěnců Všeobecné zdravotní pojišťovny (VZP) patří k systémovým i strategickým úlohám potřebným pro plánovité financování zdravotní péče pro úroveň celé ČR. Cílem příspěvku je představit východiska, koncepty a aktuální výslednice modelování počtů pojištěnců VZP ve prospěch co možná nejpřesnější predikce ročních časových řad počtů pojištěnců (podle pohlaví a věku) do roku 2022.

Data a metodika: V rámci dostupných datových zdrojů (ročenky VZP, centrální registr pojištěnců ČR) jsme analyzovali vývoj pojistného kmene VZP (celkové počty pojištěnců podle pohlaví a 5letých věkových skupin) za roky 2002-2016. V souladu s interaktivní analýzou dat byly do základního modelu zavzaty (kromě časové proměnné) navíc ještě časové řady z demografie (projekce počtů obyvatel ČR do r. 2022 (ČSÚ)) a údaje o přeregistracích pojištěnců (odchody a příchody do VZP v odpovídajících strukturách). Počty budoucích pojištěnců byly modelovány na základě pragmatických multiplikativních (ročních) dopočtů a aditivních (víceletých) regresních modelů s retrospektivně se snižujícími vahovými/diskontními faktory tzn. v podstatě adaptivním řídícím procesem.

Výsledky a závěry: V rámci jednotlivých věkových skupin se významnost obou použitých exogenních proměnných (tzn. demografické vs "přeregistrační") dosti podstatně mění (konzistentně pro obě pohlaví). Jejich simultánní zahrnutí do modelů však vysvětluje procento variability počtů pojištěnců dominantní měrou. Kvalita a predikční přesnost modelů výše ale závisí také na ročním období (měsíci), pro který je roční časová řada pojištěnců aktuálně monitorována.

Michal Černý a Miroslav Rada

Big Data a kolmogorovská složitost (a také několik reminiscencí na WSC ISI 2017)

VŠE v Praze, FIS, katedra ekonometrie, Nám. W. Churchilla 4, 130 67 Praha 3

cernym@vse.cz, miroslav.rada@vse.cz

Tento příspěvek byl motivován diskusemi ze Světového kongresu ISI 2017 o fenoménu tzv. Big Data. Z diskusí ovšem nebylo možné zjistit, co se tímto pojmem rozumí a jak je definován. Zde se budeme zabývat tzv. Narrow Big Data. Z názvu je patrné, že data mají být *velká* a *úzká*. To lze formalizovat pomocí následující definice. Velikost formalizujeme tak, že operační paměť je menší než dataset, který má být zpracován. Přesněji: data nechť jsou organizována v racionální matici A rozměru $n \times p$, kde řádky odpovídají pozorováním a sloupce dimenzi. Předpokládáme, že n je podstatně větší než p. Máme k dispozici turingovský výpočetní model s pamětí omezenou na $P(p)$ buněk, kde P je polynom, a v jedné buňce umožníme uložit racionální číslo o bitové velikosti $O(L^k \cdot \log^k n)$, kde $k \geq 1$ je konstanta a $L = \max_{ij} \text{bitsize}(A_{ij})$. Tento výpočetní model skutečně zachycuje fakt, že data A se nevejdou do paměti celá: je-li totiž $n \gg p$ (např. $n \geq \Omega(2^p)$), je možné uložit jen omezený počet datových bodů (řádků matice A) a provádět běžné operace jen s malými objekty, např. s maticemi, jejichž rozměr je omezen polynomem v p (ale nikoliv v n). Není-li možné uložit data do paměti, je třeba říci, jak se k nim přistupuje. Možností je více (např. oraculový model); v tomto příspěvku studujeme *streamový model*. Datový bod (řádek matice A) lze načíst do paměti a zpracovat; tím je ovšem bod ztracen a nadále již není přístupný. Cílem příspěvku je ukázat, že tento model vede k významným výpočetním omezením: ilustrujeme to větou, že v tomto modelu *nelze spočítat medián*. Ukážeme důkaz postavený na kolmogorovské složitosti. Cílem příspěvku je pohlédnout na počítání s Narrow Big Daty „zespodu“: na jednu stranu existuje snaha navrhovat pokročilé a sofistikované metody pro Big Data, na druhou stranu se ovšem ukazuje, že výpočetní omezení jsou natolik vážná, že nedovolují vyčíslit ani některé elementární statistiky.

Adéla Drabinová, Patrícia Martinková

Modely přidané hodnoty škol

A.D.: MFF UK, KPMS, Sokolovská 83, 186 75 Praha 8
P.M.: ÚI AV ČR, Pod Vodárenskou věží 2, 182 07, Praha 8
A.D., P.M.: PedF UK, ÚVRV, Myslíkova 7, 110 00, Praha 1

drabinova@cs.cas.cz

Míra přidané hodnoty (value-added measure) se používá k odhadu či kvantifikaci toho, jak velký pozitivní či negativní vliv má škola na studijní výsledky žáků v průběhu jejich studia. Modely pro přidanou hodnotu škol (value-added models, VAM) nám tak mohou pomoci identifikovat efektivní a kvalitní školy na základě meziročního zlepšení žáků ve standardizovaných testech. V tomto příspěvku představíme a porovnáme několik možných statistických modelů - od jednoduchých, kde efekt školy je modelován jako pevný efekt a jako znalost žáků je bráno dosažené skóre, až po složitější modely se smíšenými efekty, kde se navíc předpokládá, že znalost žáků může být latentní a je ji nutno odhadnout. Jednotlivé přístupy budou demonstrovány na reálných datech.

Zdeněk Fabián

Informace a neurčitost

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

zdenek@cs.cas.cz

Tento příspěvek pojednává o informaci a neurčitosti spojité náhodné veličiny. Půjde nám o problémy typu: Jaká je hustota informace a neurčitosti spojité náhodné veličiny? Jaké relativní množství informace a neurčitosti je obsaženo v jednom pozorování z daného rozdělení? A jaký je vlastně vztah mezi informací a neurčitostí?

Matematická statistika ani teorie informace nedávají na tyto přirozené otázky odpověď. Fisherova informace se týká parametrů parametrického rozdělení a neurčitost popsaná diferenciální entropií může být záporná.

Základem nového přístupu je idea, že namísto realizace x dané náhodné veličiny X je vhodné uvažovat hodnotu příbuzné náhodné veličiny, skórové funkce rozdělení $S(x)$, vyjadřující vliv x na určitý střed daného rozdělení. Ukážeme, že informací ve spojitých modelech je zobecněná Fisherova informace a střední neurčitost je převrácená hodnota její střední hodnoty.

Kamila Fačevicová[1], Peter Filzmoser[2], Karel Hron[1]

Alternativní přístup k analýze vícefaktorových dat

[1] Katedra matematické analýzy a aplikací matematiky, Přírodovědecká fakulta Univerzity Palackého v Olomouci,
[2] Institute of Statistics and Mathematical Methods in Economics, Vienna University of Technology

kamila.facevicova@gmail.com

Příspěvek je zaměřen na alternativní metodu analýzy vztahu mezi dvěma a více faktory. Konkrétně se budeme zabývat situací, kdy pracujeme s daty, u nichž je (namísto absolutních hodnot) relevantní jejich relativní struktura.

Právě tímto typem problémů se zabývají metody zpracování tzv. kompozičních dat [2] a analýzou vztahů mezi dvěma faktory pak teorie práce s kompozičními tabulkami [1]. Analýza kompozičních tabulek a kompozičních dat obecně je založena na jejich vhodné souřadnicové reprezentaci, která následně umožnuje použití širokého spektra standardních (nekompozičních) statistických metod. Z našeho pohledu optimální varianta této reprezentace byla představena již v rámci uplynulého Robustu 2016. Současný přispěvek na ten z roku 2016 naváže. Navržená souřadnicová reprezentace bude nejen dále zobecněna pro případ vícefaktorových kompozičních dat, ale zejména na reálném příkladu porovnána s tradičním přístupem využívajícím log-lineárních modelů.

Reference

[1] Fačevicová K, Hron K, Todorov V, Templ M (2015) *Compositional tables analysis in coordinates.* Scandinavian Journal of Statistics **43**(4), str. 962-–977.

[2] Pawlowsky-Glahn V, Egozcue JJ, Tolosana-Delgado R (2015) *Modeling and analysis of compositional data.* Wiley, Chichester.

Jana Faltýnková

Geografická profilace s využitím bayesovských metod

Masarykova univerzita, Ústav matematiky a statistiky, Kotlářská 2, 611 37 Brno

xfaltynkovaj@math.muni.cz

Geografická profilace v kriminalistice umožňuje pomáhat kriminologům nalézt kotevní bod pachatele. Prostřednictvím bayesovských metod lze do modelu implementovat priorní informace o pachateli či skupině pachatelů podobného typu. Jako výsledek získáme posteriorní funkci, která udává pravděpodobnostní rozdělení možných míst viníkova kotevního bodu. Naše metoda nabízí postup pro práci s pachateli typu „marauder“ i „commuter“ a zvažuje řadu modelů, které umožňují analytický výpočet. Porovnání výsledků našeho přístupu se známým Rossmovým modelem na reálných datech z Baltimore County ukazuje u navrženého způsobu s využitím bayesovských metod větší efektivitu při hledání pachatele, obzvláště pak pro viníky typu „commuter“, kteří za trestnými činy dojíždějí.

Reference

[1] Mohler, G. and Short, M.: *Geographic Profiling from Kinetic Models of Criminal Behavior.* SIAM Journal on Applied Mathematics. **1** (2012) 163–180.

[2] Rossmo, D.: *Geographic Profiling.* Boca Raton, Fla: CRC Press. (2000)

Eva Fišerová

Odhady kuželoseček a kvadrik a jejich přesnost

Univerzita Palackého v Olomouci, PřF, KMAAM, 17. listopadu 12, 771 46 Olomouc

eva.fiserova@upol.cz

Odhady kuželoseček a kvadrik patří mezi základní úlohy v mnoha vědních oblastech. Setkat se s nimi můžeme např. ve fyzice, astronomii, biologii, při kontrole kvality nebo zpracování obrazu. Většina optimalizačních i statistických metod je založena na minimalizaci algebraické nebo geometrické vzdálenosti. V případě algebraické vzdálenosti jednoduše dosadíme data do implicitní rovnice hledané kuželosečky/kvadriky a minimalizujeme druhou mocninu odchylek od nuly; geometrický přístup je založen na minimalizaci vzdálenosti mezi napozorovanými daty a jejich projekcí na hledanou kuželosečku/kvadriku.

V příspěvku se zaměříme na geometrický přístup při hledání odhadu kuželoseček/kvadrik pomocí lineárního regresního modelu s nelineárními omezeními na regresní parametry. Omezení na parametry, která reprezentují obecnou algebraickou rovnici hledané kuželosečky/kvadriky, lze linearizovat pomocí Taylorova rozvoje prvního řádu. Výsledný iterační algoritmus poskytuje lokálně nejlepší nestranné lineární odhady neznámých parametrů algebraické rovnice kuželosečky/kvadriky. Následně lze stanovit odhady geometrických parametrů, obsah, objem, povrch, atd., společně s odpovídajícími charakteristikami přesnosti. Výsledky nakonec zobecníme pro korelovaná data přidáním iteračního algoritmu pro odhady variančních komponent pomocí metody MINQUE.

Michal Friesl

Hokejové góly

FAV ZČU, KMA a NTIS, Univerzitní 8, 301 00 Plzeň

friesl@kma.zcu.cz

V tomto oddychovém příspěvku se podíváme na vývoj počtu gólů v průběhu utkání ledního hokeje. Na základě podrobných údajů o časech více než 100 tisíc vstřelených gólů v zápasech americké NHL v letech 1994–2015 si přiblížíme některá empirická fakta a porovnáme pak z hlediska průběhu skórování jednotlivé třetiny zápasu. Vytušíme potenciál rozdílů mezi třetinami v souvislosti se snahami o zvýšení celkového počtu vstřelených branek.

Andrej Gajdoš, Martina Hančová

Rozdelenia pravdepodobnosti odhadcov variančných parametrov vo FDSLRM

PF UPJŠ, ÚMV, Jesenná 5, 040 01, Košice 1

andrej.gajdos@student.upjs.sk

Pri praktickej analýze časových radov (ČR) je potrebné odhadovať neznáme variančné parametre. Následné dosadenie odhadov do prediktora ovplyvňuje vzniknutú tzv. empirickú predikciu, preto je vhodné poznať vlastnosti odhadcov.

V rámci FDSLRM, triedy modelov pre ČR, sme zaviedli novú zovšeobecňujúcu triedu odhadcov variančných parametrov založených na metóde najmenších štvorcov (MNŠ) — LCNE. Títo odhadcovia sú v tvare lineárnej kombinácie tzv. prirodzených odhadcov a zároveň táto naša trieda v sebe zahŕňa všetkých doteraz publikovaných MNŠ odhadcov vo FDSLRM. Hlavným výsledkom príspevku spoločne so zavedením LCNE je aj odvodenie ich momentových charakteristík ($E\{.\}$, $Cov\{.\}$, $MSE\{.\}$) a za predpokladu široko platných podmienok i odvodenie pravdepodobnostných rozdelení LCNE pre prípad normálne rozdeleného FDSLRM pozorovania.

Za prínos tiež považujeme návrh a realizáciu praktických spôsobov numerického výpočtu pravdepodobností, kvantilov a intervalov spoľahlivosti pre gama rozdielové rozdelenie, ktoré ako sme odvodili, je vo všeobecnosti rozdelením LCNE. Odvodenú teóriu a navrhnuté numerické prístupy sme aplikovali na problém výskytu záporných odhadov variančných parametrov vo FDSLRM.

Jan Hanousek

Endogenita proměnných v aplikacích z oblasti firemních financí

CERGE-EI Univerzita Karlova, Akademie věd, Politických vězňů 7, P.O.Box 882, 111 21 Praha

huskova@karlin.mff.cuni.cz

Uvažujme model

$$y_{i,t} = \boldsymbol{x}_{i,t}^{\top}\big(\boldsymbol{\beta}_i + \boldsymbol{\delta}_i I\{t \geq t_0\}\big) + e_{i,t}, \quad 1 \leq i \leq N,\ 1 \leq t \leq T, \tag{1}$$

kde parametry modelu v i-tém panelu se změní v neznámém čase t_0 z $\boldsymbol{\beta}_i$ na $\boldsymbol{\beta}_i + \boldsymbol{\delta}_i$.

Hlavním cílem přednášky bude seznámit posluchače se zkušenostmi s odhadováním případné změny v tomto modelu při měnících se vstupních parametrech pro statistiky založené na procesu

$$U_N(t) = \sum\nolimits_{i=1}^{N} \big(\widehat{\boldsymbol{\beta}}_{i,t} - \widehat{\boldsymbol{\beta}}_{i,T}\big)^{\top} \boldsymbol{C}_{i,t} \big(\widehat{\boldsymbol{\beta}}_{i,t} - \widehat{\boldsymbol{\beta}}_{i,T}\big),$$

kde $\widehat{\boldsymbol{\beta}}_{i,t}$ je odhad $\boldsymbol{\beta}$ získaný metodou nejmenších čtverců z prvních t pozorování a $\boldsymbol{C}_{i,t}$ je některá vhodná váhová matice. Dále bude prezentována aplikace na reálná finanční data.

Poděkování: Jedná se o výsledky společné práce s J. Antochem, J. Hanouskem a dalšími. Práce byla podpořena grantem GAČR P403/15/09663S.

Milan Hladík

Intervalová robustnost v lineárním programování

MFF UK, KAM, Malostranské nám. 25, 118 00 Praha 1

hladik@kam.mff.cuni.cz

Intervalová data se přirozeně vyskytují v řadě situací. Intervaly mohou representovat rozsah neznámých či odhadnutých hodnot, vznikají diskretizací, kategorizací, nebo účelně při analýze citlivosti. V našem pohledu z hlediska intervalové analýzy nepředpokládáme na intervalech žádnou distribuci. Bereme je jako množinu hodnot, které je potřeba všechny zohlednit a zpracovat. Z tohoto důvodu se intervalové metody využívají i v numerické analýze pro dosažení numericky rigorózních výpočtů.

V první části představíme intervalové počítání a základní koncepty. Zavedeme intervalové soustavy lineárních rovnic a nerovnic a vysvětlíme, co se vůbec rozumí pod pojmem řešení. Ukážeme, jak vypadá geometricky množina řešení, jak ji popsat a jak je výpočetně těžké s touto množinou pracovat. Zmíníme důležité robustní vlastnosti intervalových matic, jako například regularitu.

V druhé části se budeme věnovat vlastnímu intervalovému lineárnímu programování. Ukážeme, jaký vliv má změna dat na optimální hodnotu a optimální řešení a jak odhadovat maximální změnu těchto veličin. Zmíníme i výpočetní hledisko těchto problémů – obecně je mnoho souvisejících problémů výpočetně náročných, nicméně existují i jednodušší případy. Podíváme se na to, jaký tvar mají množiny optimálních hodnot a optimálních řešení a jak je charakterizovat. Užitečnou robustní vlastností je invariance optimální báze – za jakých podmínek to nastane a jak to výpočetně ověřit?

Reference

[1] M. Fiedler, J. Nedoma, J. Ramík, J. Rohn, and K. Zimmermann. *Linear Optimization Problems with Inexact Data*. Springer, New York, 2006.

[2] M. Hladík. Interval linear programming: A survey. In Z. A. Mann, editor, *Linear Programming – New Frontiers in Theory and Applications*, chapter 2, pages 85–120. Nova Science Publishers, New York, 2012.

Poděkování. Příspěvek byl podpořen grantem GAČR P403-18-04735S.

Vladimír Holý

Odhad parametrů spojitých procesů pomocí finančních vysokofrekvenčních dat

VŠE v Praze, náměstí Winstona Churchilla 1938/4, 130 67 Praha 3

vladimir.holy@vse.cz

Když je časová řada pozorována o vysokých frekvencích, více informace může být využito při odhadu parametrů procesu. Vysokofrekvenční data jsou ovšem kontaminována tzv. mikrostrukturním šumem, který způsobuje vychýlení odhadu volatility, pokud se šum nebere v potaz. Toto chování je studováno v literatuře zabývající se neparametrickými odhady kvadratické variance. My se zaměříme na vliv šumu při odhadu parametrických procesů. Tradiční metody ignorující šum jsou výrazně vychýlené, i pokud je rozptyl šumu poměrně malý. Ukážeme, že Wienerův proces pozorovaný v ekvidistantních časech a kontaminovaný nezávislým bílým šumem sleduje ARIMA(0, 1, 1) namísto ARIMA(0, 1, 0). Podobně, diskretizovaný a kontaminovaný Ornstein-Uhlenbeckův proces sleduje ARIMA(1, 0, 1) namísto ARIMA(1, 0, 0). Dále představíme estimátory robustní k šumu založené na metodě momentů a metodě maximální věrohodnosti pro ekvidistantní i nerovnoměrně rozložená pozorování.

Poděkování: Podpořeno z grantu IGS F4/93/2017 Vysoké školy ekonomické v Praze.

Michal Houda

Data envelopment analysis within evaluation of the efficency of firm productivity

EF JČK v Českých Budějovicích, KMI, Studentská 13, 370 05 České Budějovice

houda@ef.jcu.cz

The talk deals with the methodology of evaluating the efficiency of the use of production factors using methods of data envelopment analysis. The goal is to analyze accessible methods based on principles of data envelopment analysis in static as well as in dynamic framework, to apply these methods to accessible firm data from the food sector from several consecutive years, to compare obtained numerical results with well-known productivity indices, and to perform a statistical inference on these results.

Poděkování: Author thanks the Ministry of Education of the Czech Republic for the financial support.

Karel Hron

Vážení složek v analýze kompozičních dat

PřF UP, Katedra matematické analýzy a aplikací matematiky, 17. listopadu 12, 771 46 Olomouc

hronk@seznam.cz

Analýza D-složkových kompozičních dat, pozorování nesoucích relativní informaci, prostředky Aitchisonovy geometrie na simplexu předpokládá rovnoměrné rozdělení jako referenční míru měřitelného prostoru, jehož rozklad na D kategorií (odpovídajících kompozičním složkám) uvažujeme. Změna referenční míry pak následně odpovídá

vážení složek. Změny se projeví i v algebraticko-geometrické struktuře na simplexu, potažmo v souřadnicové reprezentaci kompozičních dat, která je nutným předpokladem použití standardních nástrojů mnohorozměrné statistiky. Volba vah závisí na podstatě studovaného problému, a může reflektovat např. přesnost měření dané složky, počet hodnot pod detekčním limitem a podobně. Speciálně se dá taktéž ukázat, že při postupném snižování vah vybraných složek bude takto geometrie prostoru v limitě odpovídat příslušné podkompozici. Teoretické výstupy budou demonstrovány na simulovaných i reálných datech.

Samuel Hudec

Priemerované diskrétne zmiešané logaritmické rozdelenia

Univerzita Mateja Bela, Fakulta prirodnych vied, Katedra matematiky, Tajovského 40,97401, Banska Bystrica

samuel.hudec@umb.sk

Príspevok je zameraný na sprehľadnenie podtriedy nezáporných diskrétnych zmiešaných rozdelení, ktoré sme nazvali priemerované zmiešané rozdelenia.Predstavená trieda rozdelení zahŕňa rozdelenia s netriviálnou pravdepodobnostnou funkciou a nekonvexným parametrickým priestorom, ktoré sa vyskytli v literatúre. Príspevok obsahuje konkrétne riešenia odhadovania parametrov v celej triede týchto rozdelení. Na záver sa simulačne preverujú asymptotické vlastnosti odhadovacích procedúr a testy dobrej zhody pri optimálne nagenerovanej vzorke. Postupy a výsledky sú naprogramované v štatistickom prostredí R a sú voľne prístupné na ďalšie modifikácie, ale aj na využitie v podobných prípadoch.

Marie Hušková

O detekci změn v panelových datech

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

huskova@karlin.mff.cuni.cz

Uvažujme model

$$y_{i,t} = \boldsymbol{x}_{i,t}^{\top}\big(\boldsymbol{\beta}_i + \boldsymbol{\delta}_i I\{t \geq t_0\}\big) + e_{i,t}, \quad 1 \leq i \leq N,\ 1 \leq t \leq T, \tag{2}$$

kde parametry modelu v i-tém panelu se změní v neznámém čase t_0 z $\boldsymbol{\beta}_i$ na $\boldsymbol{\beta}_i + \boldsymbol{\delta}_i$.

Hlavním cílem přednášky bude seznámit posluchače se zkušenostmi s odhadováním případné změny v tomto modelu při měnících se vstupních parametrech pro statistiky založené na procesu

$$U_N(t) = \sum\nolimits_{i=1}^{N} \big(\widehat{\boldsymbol{\beta}}_{i,t} - \widehat{\boldsymbol{\beta}}_{i,T}\big)^{\top} \boldsymbol{C}_{i,t} \big(\widehat{\boldsymbol{\beta}}_{i,t} - \widehat{\boldsymbol{\beta}}_{i,T}\big),$$

kde $\widehat{\boldsymbol{\beta}}_{i,t}$ je odhad $\boldsymbol{\beta}$ získaný metodou nejmenších čtverců z prvních t pozorování a $\boldsymbol{C}_{i,t}$ je některá vhodná váhová matice. Dále bude prezentována aplikace na reálná finanční data.

Poděkování: Jedná se o výsledky společné práce s J. Antochem, J. Hanouskem a dalšími. Práce byla podpořena grantem GAČR P403/15/09663S.

Martina Chvosteková

Viacnásobne použiteľné oblasti spoľahlivosti pre viacrozmernú kalibráciu

Ústav merania SAV, Dúbravská cesta 9, 841 04, Bratislava 4, Slovensko

martina.chvostekova@savba.sk

Na základe odhadnutého viacrozmerného lineárneho regresného modelu popisujúceho vzťah m-rozmernej vysvetľujúcej a q-rozmernej vysvetľovanej premennej je úlohou viacrozmernej kalibrácie stanoviť oblasť spoľahlivosti pre hodnotu vysvetľujúcej premennej prislúchajúcej k budúcemu pozorovaniu vysvetľovanej premennej. Počet budúcich pozorovaní je neohraničene veľký. Požadujeme, aby aspoň γ časť oblastí spoľahlivosti skonštruovaných k postupnosti budúcich pozorovaní pokrývala skutočnú prislúchajúcu hodnotu vysvetľujúcej premennej. Keďže oblasti spoľahlivosti sú skonštruované na základe raz odhadnutých parametrov modelu, požadované γ pokrytie je dosiahnuté s pravdepodobnosťou $1 - \alpha$. Oblasti spoľahlivosti definované pravdepodobnosťami $1 - \alpha$ a γ nazývame viacnásobne použiteľné oblasti spoľahlivosti. V príspevku porovnáme štatistické vlastnosti viacnásobne použiteľných oblastí spoľahlivosti navrhnutých z podmienky podobnej tolerančným oblastiam (pozri Mathew a Zha (1997), Mathew a kol. (1998)) a navrhnutých viacnásobne použiteľných oblastí skonštruovaných priamo z podmienky tolerančných oblastí.

Reference

[1] Mathew, T., Zha, W. (1997): Multiple use confidence regions in multivariate calibration. *Journal of the American Statistical Association* **92**, 1141–1150.

[2] Mathew, T., Sharma, M. K., Nordstrom, K. (1998): Tolerance regions and multiple - use confidence regions in multivariate calibration. *The Annals of Statistics* **26**, 1989–2013.

Poďakovanie: Táto práca bola podporená Agentúrou na podporu výskumu a vývoja APVV-15-0295 a Vedeckou grantovou agentúrou MŠVVaŠ SR a SAV VEGA 2/0011/16.

Jozef Jakubík

Nelineárna Grangerova kauzalita pomocou neurónových sietí

Ústav merania SAV Dúbravská cesta 9, 841 04 Bratislava 4, Slovensko

jozef.jakubik.jefo@gmail.com

Pri skúmaní vzťahu medzi dvoma časovými radmi (vplýva časový rad X na časový rad Y?) sa často využíva Grangerova kauzalita (GK). GK vyšetruje, nakoľko pridanie informácie z časového radu Y pomôže pri predikcii časového radu X. Ako sme ukázali v článku, GK funguje dobre najmä v prípadoch keď sú časové rady AR proces. V prípade zložitejších systémov, napríklad Henónov:

$$x_1(n+1) = 1.4 - x_1^2(n) + 0.3x_2(n)$$
$$x_2(n+1) = x_1(n)$$

už neposkytuje uspokojujúce výsledky. V posledných rokoch bolo navrhnutých viacero rozšírení GK alebo iných metód (porovnanie niektorých z nich je možné nájsť v článku). Tieto metódy lokálne aproximujú nelineárny systém, predikciu zakladajú na dátach z minulosti, ktoré boli podobné súčasnosti.

Pri aproximácii neznámych funkcií sa s rozmachom výpočtových možností počítačov začali hojne využívať neurónové siete. Preto nami navrhnutá metóda využíva na predikciu neurónové siete. V porovnaní s niektorými metódami (ktoré sa vyhodnocujú „od oka“) dokážeme sformulovať test, s ktorého pomocou môžeme exaktne vyhodnotiť príspevok časového radu Y k predikcii časového radu X.

Poďakovanie: Táto práca bola podporená Agentúrou na podporu výskumu a vývoja APVV-15-0295 a Vedeckou grantovou agentúrou MŠVVaŠ SR a SAV VEGA 2/0011/16.

Josef Janák

Odhady parametrů v rovnici stochastického oscilátoru

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

janak@karlin.mff.cuni.cz

Uvažujme následující vlnovou rovnici

$$\begin{aligned}
\frac{\partial^2 u}{\partial t^2}(t,\xi) &= b\Delta u(t,\xi) - 2a\frac{\partial u}{\partial t}(t,\xi) + Q^{\frac{1}{2}}\dot{B}(t,\xi), \quad (t,\xi) \in \mathbb{R}_+ \times D,\\
u(0,\xi) &= u_1(\xi), \quad \xi \in D,\\
\frac{\partial u}{\partial t}(0,\xi) &= u_2(\xi), \quad \xi \in D,\\
u(t,\xi) &= 0, \quad (t,\xi) \in \mathbb{R}_+ \times \partial D,
\end{aligned}$$

kde $D \subset \mathbb{R}^d$ je otevřená omezená množina s hladkou hranicí, $a > 0$ a $b > 0$ jsou neznámé parametry, $Q^{\frac{1}{2}}$ je pozitivní nukleární operátor v $L^2(D)$ a $\dot{B}(t,\xi)$ je formální časová derivace Brownova pohybu závislého na prostorové proměnné.

V příspěvku pořídíme silně konzistentní odhady parametrů a a b, a to jak na základě pozorování celé trajektorie procesu $(X(t) = (u(t,\cdot), \frac{\partial u}{\partial t}(t,\cdot))^\top, 0 \le t \le T)$, tak na základě pozorování přes "pozorovací okno" $\langle X(t), z \rangle_V$ (tj. známe–li například pouze určité souřadnice). Ukážeme, že tyto odhady mají asymptoticky normální rozdělení a na příkladu kmitání tyče představíme jejich implementaci v praxi.

Daniela Jarušková

Návrh experimentu v jednom problému nelineární regrese s náhodnými parametry

FSv ČVUT, katedra matematiky, Thákurova 7, 166 29 Praha 6

daniela.jaruskova@cvut.cz

Studovaný problém spočívá v navržení experimentu pro jeden speciální případ nelineární regrese pocházející z termofyziky. Předpokládá se, že regresní funkce obsahuje kromě odhadovaných parametrů ještě další parametry, které jsou náhodné a o kterých předpokládáme, že je jejich rozdělení známé. K témuž problému můžeme přistupovat frekvenčně či bayesovsky, ale i v rámci frekvenčního přístupu můžeme postupovat při navrhování a vyhodnocování experimentu různě. Příspěvek ukazuje, že "optimální řešení"je velmi silně ovlivněno tím, jakým způsobem se na problém díváme.

Reference

[1] Jarušková D., Kučerová A.: Estimation of thermophysical parameters from the point of view of non-linear regression with random parameters, International Journal of Heat and Mass Transfer 106, 2017, 135-141.

Čeněk Jirsák

Optimalizace řízení redundantního systému k z n pomocí metody simulovaného žíhání

FP TUL, KAP, Univerzitní náměstí 1410/1, 461 17 Liberec

cenek.jirsak@tul.cz

Simulované žíhání je stochastický optimalizační algoritmus založený na Markov chain Monte Carlo metodách. V našem případě je aplikován na úlohu řízení skupiny paralelně pracujících a spojitě zastarávajících komponent. Konkrétně nás zajímá redundantní systém komponent, který je systém funkční, pokud funguje alespoň k z jeho n komponent.

Úloha je inspirována technickou praxí, kde zatím převládají jednoduché modely (dvoustavové, případně diskrétní markovské řetězce). Pro spojitý popis stavu komponenty jsou známy výsledky v uzavřeném tvaru jen pro jednoduché případy. Pro složitější situace používáme numerické algoritmy.

Karel Kadlec

Ergodic Control for stochastic equations in Hilbert spaces with Lévy noise

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

kadlec@karlin.mff.cuni.cz

In this contribution, controlled linear stochastic evolution equations driven by Lévy processes are presented in the Hilbert space setting. The control operator may be unbounded which makes the results obtained in the abstract setting applicable to parabolic SPDEs with boundary or point control. In the first part, some examples, as various parabolic type SPDEs with point or boundary control, are introduced. The second part contains some preliminary technical results, notably a version of Itô formula which is applicable to weak/mild solutions of controlled equations. In the last part, the ergodic control problem is solved: The feedback form of the optimal control and the formula for the optimal cost are found.

Jan Klaschka

Exaktní intervalové odhady prevalence vzácnějších vrozených vad. Rozpory s testy a jak je řešit.

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

klaschka@cs.cas.cz

V rámci grantového projektu hodnotíme s týmem pracovníků Thomayerovy nemocnice v Praze a ÚI AV ČR vliv prenatální diagnostiky na výskyt některých vrozených vývojových vad u dětí narozených v ČR. Některé typy vad jsou natolik frekventované, že nelze mít vážné námitky, jsou-li pro intervalové odhady a testy týkající se jejich prevalence použity asymptotické statistické metody. Jiné typy vad jsou však o mnoho vzácnější, a pak musíme sáhnout po metodách exaktních. Některé z těchto metod ovšem vykazují určité patologické rysy – konfidenční množiny nejsou intervaly nebo sice intervaly jsou, ale vzájemně si odporují s příslušnými testy. Nechcemeli analýzu prevalencí vrozených vad zatěžovat logickými rozpory mezi inferencemi, je pak řešením problému

drobná úprava testu založená na tzv. unimodalizaci funkce p-hodnoty (p-value function, též evidence function a další názvy v literatuře, viz [2].) Unimodalizace spočívá v nahrazení původní funkce f funkcí g, která je unimodální, majorizuje f a je mezi funkcemi s těmito vlastnostmi stejnoměrně nejmenší. Programy v R z dílny autora a maďarského kolegy J. Reiczigela, které počítají upravené p-hodnoty pro Sterneho [3] a Blakerovu [1] metodu v kombinaci s binomickým a Poissonovým rozdělením, dílem jsou a dílem brzy budou volně dostupné na internetu.

Reference

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

[2] Hirji K. F. (2000). *Exact Analysis of Discrete Data.* Chapman and Hall/CRC, New York.

[3] Sterne T. E. (1954). *Some remarks on confidence or fiducial limits.* Biometrika **41**, 275–278.

Poděkování: Práce byla podpořena grantem AZV 17-29622A.

Jana Klicnarová

Modifikace Randlesových nadrovin

EF JČU v Českých Budějovicích, KMA, Studentská 13, České Budějovice

klicnarova@ef.jcu.cz

V mnoha ekonomických i jiných aplikacích je nutné využívat vícerozměrné neparametrické testy. Jedním z možných přístupů k těmto testům jsou metody založené na konceptu Randlesových nadrovin. Randles ve svém článku (Randles (89)) představil jednovýběrový test polohy založený na počtu nadrovin procházejících středem rozdělení a oddělujících body pozorování. Jeho výsledky byly později zobecněny a vylepšeny různými autory a byly získány dvou i vícevýběrové testy parametrů polohy, testy nezávislosti aj.

Cílem příspěvku je připomenout tyto výsledky a ukázat další možné modifikace těchto přístupů, včetně jejich možných výhod a nevýhod. Jedná se o předběžné výsledky dosažené ve spolupráci s Miroslavem Šimanem a Davym Paindaveinem.

Poděkování: Příspěvek vznikl částečně za podpory GA ČR, projekt č. GA17-07384S. Autorka také děkuje za finanční podporu MŠMT ČR.

Kateřina Koňasová

Stochastická rekonstrukce pro nehomogenní bodové procesy

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

konasova.k@seznam.cz

Při zkoumání vlastností bodového vzorku se zpravidla snažíme vybrat parametrický model, který by co nejlépe odpovídal analyzovaným datům. V dalším kroku odhadujeme parametry modelu a testujeme shodu s pozorovanými daty. Možnost simulovat z odhadnutého modelu je v prostorové statistice klíčová, neboť nám umožňuje získat představu o chování pozorovaného vzorku mimo pozorovací okno či zkoumat variabilitu empirických odhadů popisných charakteristik. Pro konkrétní data však může být odpovídající teoretický model komplikovaný a práce s ním tudíž obtížná. V některých případech dokonce ani nemusíme být schopni odhadnout parametry modelu.

Stochastická rekonstrukce je iterační procedura, která nám dovoluje generovat bodové vzorky s předepsanými hodnotami popisných charakteristik. Celý postup navíc nevyžaduje volbu konkrétního teoretického modelu pro pozorovaná data. Algoritmus má svůj původ ve statistické fyzice, dnes je oblíben především v oblasti biologie a ekologie.

Doposud byla stochastické rekonstrukce uvažována pouze ve spojitosti se stacionárními bodovými procesy. Naším cílem je nastínit možnosti rozšíření algoritmu z článku [1] pro třídu bodových procesů s vlastností SOIRS. Zaměříme se na využití nehomogenní varianty K a J-funkce a na zavedení kritéria pro hodnocení kvality výstupu této procedury.

Poděkování: Tato práce je podporována Grantovou agenturou Univerzity Karlovy, projekt č. 472217.

Literatura

[1] A. Tscheschel, D. Stoyan (2006). Statistical reconstruction of random point patterns. Computational Statistics & Data Analysis, **51**, 859-871.

Kateřina Konečná

Metody odhadu vyhlazovacích parametrů Priestley-Chaova odhadu podmíněné hustoty

FS VUT v Brně, Ústav matematiky a deskriptivní geometrie, Žižkova 17, 602 00, Brno
Ústav matematiky a statistiky, Přírodovědecká fakulta, Masarykova univerzita, Kotlářská 2, 611 37, Brno

konecna.k@fce.vutbr.cz

Jádrové vyhlazování je v praxi často využívanou neparametrickou metodou, příspěvek je věnován jádrovým odhadům podmíněné hustoty. Nejčastěji používaným odhadem je Nadaraya-Watsonův odhad, my se zaměříme na nový typ odhadu - Priestley-Chaův odhad podmíněné hustoty, který vychází z Priestley-Chaova odhadu regresní funkce.
Kromě statistických vlastností odhadu budou uvedeny lokální a globální míry kvality odhadu, které jsou klíčovými charakteristikami pro odhad vyhlazovacích parametrů. Zásadní vliv na kvalitu jádrového odhadu mají právě vyhlazovací parametry, jejichž optimální šířky ovšem závisí na skutečné podmíněné a marginální hustotě. V případě reálných dat tyto charakteristiky nejsou obvykle známy, pro praktické využití je potřeba navrhnout metody pro odhad šířek vyhlazovacích parametrů. Vhodnost navržených metod bude porovnána pomocí simulační studie.

Miloš Kopa, Thierry Post

Portfolio optimization with DARA stochastic dominance constraints

M.K.: MFF UK, KPMS, Sokolovská 83, 186 75 Praha 8
T.P.: Graduate School of Business of Nazarbayev University, Astana, Kazachstán

kopa@karlin.mff.cuni.cz

An optimization method is developed for constructing investment portfolios which stochastically dominate a given benchmark for all decreasing absolute risk-averse investors, using Quadratic Programming. The method is applied to standard data sets of historical returns of equity price reversal and momentum portfolios. The proposed optimization method improves upon the performance of Mean-Variance optimization by tens to hundreds of basis points per annum, for low to medium risk levels. The improvements critically depend on imposing the complex condition of Decreasing Absolute Risk Aversion in addition to the simpler conditions of global risk aversion and decreasing risk aversion.

Monika Kroupová, Ivana Horová, Jan Koláček

Jádrové odhady gradientu regresní funkce

Masarykova univerzita, Ústav matematiky a statistiky, Kotlářská 2, 611 37 Brno

379157@mail.muni.cz

Jedním z nejdůležitějších faktorů ve vícerozměrném jádrovém odhadu gradientu regresní funkce je volba vyhlazovací matice. Tato volba je zvláště důležitá, protože ovlivňuje množství vyhlazování v daném směru. V tomto příspěvku se omezíme pouze na diagonální vyhlazovací matice. Použitá metoda je založená na vyváženém vztahu mezi rozptylem a vychýlením odhadu gradientu regresní funkce. Představíme simulační studii porovnávající navrženou metodu s metodou křížového ověření. V závěru příspěvku aplikujeme metodu také na reálná data.

Vít Kubelka

Linear filtering of general gaussian processes

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

kubelka@karlin.mff.cuni.cz

Some earlier results on Kalman-Bucy filter for signals given by general gaussian processes will be recalled and possible extensions to infinite dimensions will be discussed.

Petr Lachout

Stochastické optimalizační schéma s hodinkami

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

Petr.Lachout@mff.cuni.cz

V tomto příspěvku bude představeno stochastické optimalizační schéma, v němž rozhodovací body budou nejdříve neaktivní a teprve po nastání nějaké události se postupně stanou aktivními. Úkolem je optimalizovat výnos náhodného procesu, který řídíme prostřednictvím našich rozhodnutí v aktivních rozhodovacích bodech.
V přednášce schéma popíšeme a pokusíme se nabídnout postup hledání optimálního řízení.

Lívia Leššová, Michaela Koščová

Opakované parciálne sumácie

FMFI UKo, Katedra aplikovanej matematiky a štatistiky, Mlynská dolina, 842 48 Bratislava

livia.lessova@fmph.uniba.sk

Príspevok bude hovoriť o opakovaných parciálnych sumáciách diskrétnych rozdelení pravdepodobnosti. Jednorozmerné parciálne sumácie definujeme ako

$$P_x^{(1)} = \sum_{j=x}^{\infty} g(j) P_j^*, \qquad x = 0, 1, 2, ..., \tag{3}$$

kde $\{P_j^*\}_{j=0}^{\infty}$ je diskrétne rozdelenie pravdepodobnosti, tzv. rodič, $\{P_x^{(1)}\}_{x=0}^{\infty}$ je tiež diskrétne rozdelenie pravdepodobnosti, tzv. potomok, a $g(j)$ je reálna funkcia, ktorá charakterizuje parciálnu sumáciu.

Dosaďme potomka $\{P_x^{(1)}\}_{x=0}^{\infty}$, ktorý vznikol parciálnou sumáciou (3), za rodiča ďalšej generácie. Dostaneme potomka $\{P_x^{(2)}\}_{x=0}^{\infty}$. Tento postup opakujme ďalej ($\{P_x^{(n)}\}_{x=0}^{\infty}$ je rodičom potomka $\{P_x^{(n+1)}\}_{x=0}^{\infty}$). Otázkou je, či existuje pre nejakého konkrétneho rodiča $\{P_x^*\}_{x=0}^{\infty}$ a konkrétnu funkciu $g(j)$ limitné rozdelenie takýchto opakovaných sumácií, teda rozdelenie $\{P_x^{(\infty)}\}_{x=0}^{\infty}$. Pre $g(j) = c$ a pre širokú triedu rodičovských rozdelení $\{P_x^*\}_{x=0}^{\infty}$ bolo v [2] dokázané, že limitné rozdelenie existuje a je ním geometrické rozdelenie.

Pre niektoré rozdelenia definované na konečnej množine sa limitné rozdelenie dá nájsť pomocou mocninovej metódy (Power Method, pozri [1]), ktorá bola navrhnutá na hľadanie vlastných čísel matíc pomocou iteračného algoritmu.

Dvojrozmerné parciálne sumácie sú definované vzťahom

$$P_{x,y} = \sum_{j=x}^{\infty} \sum_{k=y}^{\infty} g(j,k) P_{j,k}^*, \qquad x, y = 0, 1, 2, ...,$$

kde $\{P_{j,k}^*\}_{j,k=0}^{\infty}$ je rodič v dvojrozmernom prípade, $\{P_{x,y}\}_{x,y=0}^{\infty}$ je potomok a $g(j,k)$ je reálna funkcia, ktorá tiež charakterizuje parciálnu sumáciu. Ak sú opakované sumácie robené analogicky ako v jednorozmernom prípade, dá sa na určenie limitného rozdelenia za podobných podmienok ako v jednorozmernom prípade tiež využiť mocninová metóda.

Literatura

[1] Anton H., Rorres Ch. (2010). *Elementary Linear Algebra: Applications Version,* Wiley, 872-883

[2] Mačutek J. (2006). *A Limit Property of the Geometric Distribution,* Theory of Probability and its Applications, 50(2), 316–319

Poďakovanie: Podporené grantom VEGA 2/0047/15.

Ján Mačutek

O tom, čo robí štatistik s lingvistickými dátami a čo robia lingvistické dáta so štatistikom

FMFI UKo, Katedra aplikovanej matematiky a štatistiky, Mlynská dolina, 84225 Bratislava, Slovensko

jmacutek@yahoo.com

V príspevku budú prezentované ciele a metódy kvantitatívnej lingvistiky (pozri [1]) a problémy, s ktorými sa štatistik stretáva pri analýze a modelovaní lingvistických dát. V prvej časti predstavíme niektoré známe matematické modely jazykových zákonov (napr. Zipfov a Menzerathov-Altmannov zákon). Ukážeme, ako môžu štatistické

metódy (aplikované buď priamo na dáta, alebo na parametre modelov) pomôcť pri riešení tak klasických lingvistických problémov, ako sú napríklad automatická klasifikácia textov alebo rozhodovanie o autorstve textov, ako aj v oblastiach presahujúcich čisto lingvistický výskum (napr. detekcia kognitívneho regresu spôsobeného Alzheimerovou chorobou alebo identifikácia ľudí so samovražednými sklonmi). Ak je prvá časť príspevku opisom výhod, ktoré prináša štatistika lingvistovi, druhá časť vidí veci z opačného uhla pohľadu. Zmienime sa o tom, ako sa môžu lingvistické dáta stať inšpiráciou pre štatistický výskum (spomenieme odhaľovanie vzťahov medzi rôznymi rozdeleniami pravdepodobnosti a problémy súvisiace s vyhodnotením miery zhody medzi modelom a dátami).

Poďakovanie: Podporované grantom VEGA 2/0047/15.

Literatura

[1] Köhler, R., Altmann, G., Piotrowski, R.G. (eds.) (2005). *Quantitative Linguistics. An International Handbook.* Berlin, New York: de Gruyter.

Miroslav Magát

Zhlukovanie časových radov s chýbajúcimi hodnotami

FPV UKF, KM, Tr. A. Hlinku 1, 949 74 Nitra

miro.magat@gmail.com

Reálne dáta vo forme časových radov, ktoré sú získavané meraním, často obsahujú veľa chýbajúcich pozorovaní (príčinou môže byť výpadok napájania meracieho prístroja alebo nameraná hodnota mimo rozsahu meracieho prístroja). Analyzovanie priebehu takýchto časových radov štandarnými metódami po odstránení chýbajúcich hodnôt často nie je možné z dôvodu malého počtu pozorovaní. Závažnejsím problémom však je skreslenie priebehu pôvodného časového radu, ak z neho odstránime chýbajúce hodnoty (dôjde k zanedbaniu časovej súvislosti pozorovaní).

Príspevok prezentuje problém nájdenia podobných časových radov využitím metód fuzzy zhlukovej analýzy, ktoré dokážu pracovať aj s časovými radmi obsahujúcimi veľa chýbajúcich pozorovaní. Dáta pochádzajú z reálnych meraní vo vybraných mestách Slovenska.

Tomáš Masák

Robustní analýza hlavních komponent

Technische Universität München, Lehrstuhl für Mathematische Statistik,
Parkring 13, 85748 Garching b. München

tom.masak@gmail.com

Výchozí předpoklad analýzy hlavních komponent (PCA) zní, že danou matici X lze aditivně rozložit jako $X = L+N$, kde L je matice nízké hodnosti (signál) a N je šum. Robustní PCA zobecňuje tento předpoklad uvažováním rozkladu $X = L + S + N$, kde S je řídká matice s relativně malým počtem nenulových prvků, jejichž pozice jsou neznámé. Aplikace robustní PCA přesa- hují pouhou robustifikaci populární metody analýzy dat. Pomocí uvedeného rozkladu lze například na daném videu oddělovat pohyblivé objekty (řídká matice S) od pozadí (matice nízké hodnosti L) při drobných změnách tohoto pozadí případně variaci snímání celkového obrazu (matice šumu N).

Hledání rozkladu $X = L+S+N$ je špatně podmíněná úloha. Přesto existují postupy, které dokáží za mírných předpokladů nalézt matice L, S a N pouze na základě znalosti X. Různé přístupy se liší jak v optimalizační formulaci úlohy, tak v samotných algoritmech pro minimalizaci zvolené ztrátové funkce. Jelikož přirozená formulace úlohy vede na NP-těžký problém, hlavní výzvy spočívají v navržení efektivních (uvolněných) formulací úlohy a příslušných algoritmů pracujících s minimem časově úsporných iterací.

Námi zvolená optimalizační formulace vede na algoritmus iterativně vážených nejmenších čtverců. Díky nekonvenčnímu způsobu převažování vykazuje náš algoritmus vyšší než lineární rychlost konvergence, přičemž jednotlivé iterace jsou časově i paměťově šetrné. Experimentální výsledky dále ukazují, že algoritmus je efektivnější než nejmodernější přístupy, co se týče škálovatelnosti vzhledem k rostoucí hodnosti matice L a klesající řídkosti matice S, tedy zvyšující se složitosti dané úlohy.

Markéta Matulová, Michal Koláček

Využití hybridní metody vícekriteriálního rozhodování za nejistoty k vytvoření rozhodovacího rámce pro výběr lokace

ESF MU, Lipová 41a, 602 00 Brno

marketa.matulova@econ.muni.cz

V příspěvku představujeme užitečný nástroj pro vícekriteriální rozhodování za podmínek nejistoty. Navrhovaná metoda zahrnuje několik kroků: zjištění možných alternativ, stanovení kritérií rozhodování, aplikace fuzzy přístupu pro kvantifikaci hodnot kritérií a vyhodnocení alternativ hybridní metodou kombinující fuzzy TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) a fuzzy AHP (Analytický Hierarchický Proces). Navrženou metodu aplikujeme v případové studii ke konstrukci rozhodovacího rámce pro výběr nejvhodnějšího místa pro uspořádání konference.

Stanislav Nagy

O symetrii viacrozmerných náhodných veličín

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

nagy@karlin.mff.cuni.cz

Na rozdiel od distribúcií na reálnej osi, vo viacrozmerných priestoroch neexistuje jednoznačne prijímaná definícia symetrie rozdelenia. Niekoľko rôznych prístupov kategorizujú Zuo a Serfling (2000), z čoho neskôr vychádza Serfling (2006) pri uvádzaní týchto definícií do štatistickej literatúry. V príspevku preskúmame niektoré tvrdenia Zua a Serflinga (2000) a ukážeme, že najzaujímavejšie dôkazy v ich článku nie sú úplné. Pri ďalšom skúmaní týchto problémov narazíme na neznámy dôkaz Funkovej charakterizácie symetrie konvexných telies — problému, formulovaného v roku 1913, ktorý bol vyriešený až v roku 1970. Pokiaľ vieme, jedná sa o prvý elementárny dôkaz tohto významného tvrdenia v literatúre.

Literatura

[1] Serfling, R. (2006). Multivariate symmetry and asymmetry. *Encyclopedia of Statistical Sciences, Second Edition*, 8:5338–5345.

[2] Zuo, Y. and Serfling, R. (2000). On the performance of some robust nonparametric location measures relative to a general notion of multivariate symmetry. *J. Stat. Plan. Inference*, 84(1-2):55–79.

Radim Navrátil

Analýza nákupního košíku - historie a současnost

PřF MU, ÚMS, Kotlářská 2, 611 37 Brno

navratil@math.muni.cz

Analýza nákupního košíku se snaží nalézt vzory nákupního chování ve formě asociačních či sekvenčních pravidel. Tato pravidla mohou být dále využívána obchody v doporučovacích systémech, při sestavování nabídkových balíčků, k určování obsahu propagačních katalogů a především při zacílení marketingových kampaní na stávající zákazníky. Ačkoli je z názvu metody patrné, že se původně týkala především obchodních řetězců, dnes má své uplatnění ve všech společnostech nabízejících více různých výrobků nebo služeb (bankovnictví, pojišťovnictví, průmysl,...).

Cílem příspěvku je přiblížit tuto problematiku matematickému publiku a upozornit na některé nedostatky a chyby, které se objevují v literatuře. Dále ukážeme některá její zobecnění, která se v současné době používají a ilustrujeme si je na názorných příkladech.

Robert Navrátil

Maximum volatility portfolio

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

navratil.r7@gmail.com

We shall study a maximal volatility portfolio that treats all assets in a symmetric way and a related option contract. To preserve the symmetry, we need numeraire that treats all assets symmetrically. We choose a market index with equal weights. In case of two assets, we focus on a variation of a passport option on the portfolio. The optimal strategy for the investor is the mentioned maximal volatility portfolio. We extend the known optimal strategy for the option to a richer class of convex payoff functions. We also show a modification of the optimal strategy for maximizing the probability of ending above or at a desired level. Finally, the model is extended to an arbitrary number of assets and properties of this model are discused.

Petr Novák

Odhad spolehlivosti kolejových obvodů z nekompletních cenzorovaných dat

FIT ČVUT, KAM, Thákurova 9, 160 00 Praha 6

petr.novak@fit.cvut.cz

Na základě záznamů z deseti let provozu se snažíme odhadnout rozdělení doby do poruchy jednotlivých komponent kolejových obvodů, detekujících přítomnost vlaků na traťových úsecích v ČR. Vzhledem k nekompletní evidenci poruch vycházíme z metod analýzy přežití pro cenzorovaná data. Využíváme také bayesovského přístupu pro případy, kdy porucha byla sice zaznamenána, ale nebylo možné ji jednoznačně přiřadit ke konkrétnímu zařízení či stanici. V příspěvku porovnáváme parametrické i neparametrické odhady a zkoumáme jejich vlastnosti.

Daniela Novotná

Central limit theorem for functionals of Gibbs particle processes

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

novotna@karlin.mff.cuni.cz

Two known techniques from the point process theory in the Euclidean space $\mathbb{R}^d$ are extended to the space of compact sets on $\mathbb{R}^d$ equipped by the Hausdorff metric. First, conditions for the existence of the stationary Gibbs point process with given conditional intensity have been simplified recently. Secondly, the Malliavin-Stein calculus was applied to the estimation of the Wasserstein distance between the Gibbs input and standard Gaussian distribution. We transform these theories to the space of compact sets and use them to derive a central limit theorem for functionals of a planar Gibbs segment process.

Zbyněk Pawlas

Náhodné mozaiky

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

pawlas@karlin.mff.cuni.cz

Náhodné mozaiky patří mezi nejstudovanější modely stochastické geometrie. Mozaikou v oblasti $S \subseteq \mathbb{R}^d$ rozumíme nejvýše spočetnou kolekci kompaktních množin (tzv. buněk) s neprázdnými vnitřky takovou, že jejich sjednocení vyplní celou množinu S, vnitřky buněk jsou po dvou disjunktní a každou omezenou množinu protíná pouze konečný počet buněk. Existuje řada způsobů, jak zkonstruovat modely pro náhodné mozaiky. V přednášce představíme některé z nich a uvedeme, jaké matematické problémy nabízejí. Zmíníme také statistické úlohy spojené se studiem těchto modelů.

Jan Picek

Odhady návratových hodnot klimatologických dat

FP TUL, KAP, Studentská 2, 461 17 Liberec

jan.picek@tul.cz

Příspěvek se věnuje problematice odhadů návratových hodnot pro klimatologická data. Zdánlivě jednoduchá statistická úloha, tj. odhad vysokých kvantilů, má však řadu úskalí. Zajímají nás především vysoké hodnoty kvantilů proměnných jako jsou srážky, teploty, ale na druhé straně máme k dispozici obvykle relativně krátkou dobu pozorování a je tedy otázkou, zdali běžně používané modely založené na asymptotickém rozdělení jsou vhodné. Příspěvek proto diskutuje různou volbu modelu a odhadu parametrů a na reálných datech ukazuje vliv této volby na konečný výsledek.

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

Ondřej Pokora, Jan Koláček

Analýza funkcionálních dat záznamů vyvolaných potenciálů ve sluchové dráze

PřF MU, Ústav matematiky a statistiky, Kotlářská 2, 611 37 Brno

pokora@math.muni.cz, kolacek@math.muni.cz

Vyvolané potenciály (EP) odráží neuronální aktivitu a jsou často využívány ke studiu smyslového vnímání. Příspěvek se zabývá statistickou analýzou záznamů EPI (evoked potential integral) vyvolaných potenciálů ve sluchové dráze v modelu tinnitu u krys. Ukážeme výsledky dosažené technikami analýzy funkcionálních dat, např. analýzou vzdáleností funkcionálních dat nebo funkcionální analýzou hlavních komponent.

Julie Rendlová, Karel Hron, Ondřej Vencálek, David Friedecký, Alžběta Gardlo

Bayesovský přístup k t-testům v kompoziční analýze metabolomických dat

J.R., K.H., O.V.: Katedra matematické analýzy a aplikací matematiky, Přírodovědecká fakulta, Univerzita Palackého, Olomouc, 17. listopadu 12, 771 46 Olomouc
D.F., A.G.: Oddělení klinické biochemie, Fakultní nemocnice Olomouc, I.P. Pavlova 6, 775 20 Olomouc
D.F., A.G.: Laboratoř metabolomiky, Ústav molekulární a translační medicíny, Univerzita Palackého, Olomouc, Hněvotínská 5, 779 00 Olomouc

julie.rendlova@gmail.com

Cílená či necílená metabolomika klinických vzorků představuje v poslední době slibnou cestu, jak najít nové biomarkery umožňující lepší predikci vybraných onemocnění. Nezbytnou součástí metabolomických experimentů je aplikace základních jednorozměrných a stejně tak pokročilých vícerozměrných statistických metod pro hledání nejvíce diskriminujících metabolitů obvykle mezi skupinou zdravých jedinců a nemocných. Po pre-procesingu metabolomických dat jsou rozdíly mezi pacienty a kontrolami často vyhodnocovány pomocí t-testů nebo Wilcoxonových testů. Cílem příspěvku je navrhnout bayesovskou verzi tohoto tradičního přístupu.

Metody bayesovské inference upravují apriorní pravděpodobnosti všech možných hypotéz nebo hodnot parametrů v souladu s evidencí v datech až je dosaženo aposteriorního rozdělení [1]. Bayesovský t-test předpokládá apriori místo normálního rozdělení dat raději t-rozdělení s těžšími chvosty. Díky tomu jde o přirozeně robustní metodu [2]. Navíc narozdíl od klasických statistických metod bayesovská statistika nepotřebuje při vhodné volbě apriorního rozdělení korekce na p-hodnoty u vícenásobného simultánního testování. Při vyhodnocování markerů je možné pracovat s průměrnými hodnotami aposteriorních hustot nebo s komplexnější informací zohledňující celá aposteriorní rozdělení.

Vzhledem k tomu, že je metabolom v libovolném biologickém materiálu složený z mnoha metabolitů, by metabolomická měření měla být chápána jako kompoziční data. Jejich analýza by tedy měla stavět raději na relativní struktuře metabolitů než na absolutních hodnotách spektroskopických měření. Kompoziční data se řídí Aitchisonovou geometrií namísto euklidovské geometrie v reálném prostoru [3], takže porozumění základním principům této geometrie a práce v centrovaných logratio koeficientech je nezbytná pro jakoukoliv statistickou analýzu včetně bayesovské.

Teoretická část příspěvku je ilustrovaná na analýze krevních skvrn pacientů s dědičným metabolomickým onemocněním (deficit acyl-CoA dehydrogenázy se středně dlouhým řetězcem, MCADD). Jsou připojeny dvě simulace pro porovnání stability navržené metody a tradičních t-testů v případě úbytku pozorování a při výskytu systematické chyby při měření vzorků.

Literatura

[1] Kruschke, J. (2011), Doing Bayesian Data Analysis, New York: Academic Press.

[2] Kruschke, J. K. (2012), Bayesian Estimation Supersedes the T Test, Journal of Experimental Psychology: General, 142 (2), 573-603.

[3] Pawlowsky-Glahn, V., Egozcue, J.J., Tolosana-Delgado, R. (2015), Modeling and Analysis of Compositional Data, Wiley, Chichester.

Tomáš Rusý

An asset – liability management stochastic program of a leasing company

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

rusy@karlin.mff.cuni.cz

Our talk/poster present a multi–stage stochastic program of an asset–liability management problem of a leasing company. We analyse model results as well as introduce a stress–testing methodology suited for financial applications. First, we show formulation of the business model of such a company with three various risk constraints,

namely the chance constraint, the Value–at–Risk constraint and the conditional Value–at–Risk constraint along with the second–order stochastic dominance constraint, which are applied to the model to control risk of the optimal strategy. We also talk about the structure and the generation process of the scenarios. To capture the evolution of interest rates the Hull–White model is used. Thereafter, results of the model and the effect of the risk constraints on the optimal decisions are thoroughly investigated. In the final part, the performance of the optimal solutions of the problems for unconsidered and unfavourable crisis scenarios is inspected. The methodology of a stress test we used was proposed in such a way that it answers typical questions asked by asset–liability managers.

Veronika Římalová, Eva Fišerová, Alessandra Menafoglio

Analýza prostorově závislých funkcionálních dat

V.Ř., E.F.: Katedra matematické analýzy a aplikací matematiky, Přírodovědecká fakulta Univerzity Palackého v Olomouci, 17. listopadu 12, 771 46, Olomouc
A.M.: MOX - Department of Mathematics, Politecnico di Milano, Piazza Leonardo da Vinci 32, 201 33, Milan, Italy

veronikarimalova@seznam.cz, eva.fiserova@upol.cz,
alessandra.menafoglio@polimi.it

Analýza funkcionálních dat (FDA) je soubor metod pro statistickou analýzu složitých datových struktur, jako jsou např. křivky či plochy. V reálných úlohách máme k dispozici množinu diskrétních pozorování, na jejímž základě je třeba vytvořit funkcionální pozorování, s nimiž je pak možné dále pracovat. Jednou z nejčastěji užívaných technik je použití splajnového vyhlazování se zapojením vyhlazovacího parametru, jehož nejvhodnější hodnotu lze určit např. pomocí zobecněné kros-validace.
V případě prostorových dat se často setkáme s jejich vzájemnou závislostí. Pro většinu metod z FDA je však zásadní předpoklad nezávislosti jednotlivých funkcionálních pozorování a jeho nesplnění může vést k nesprávným výsledkům. Případnou prostorovou závislost je potřeba vzít v úvahu a přizpůsobit jí použité statistické metody. Vyšetření prostorové závislosti v datech je prováděno pomocí funkcionální verze variogramu, techniky známé z klasické geostatistiky. Zkoumaná problematika bude demonstrována na reálných geologických datech.

Nikola Štefelová, Andreas Alfons, Javier Palarea-Albaladejo, Peter Filzmoser, Karel Hron

Robustní regrese s kompozičními vysvětlujícími proměnnými s odlehlostí na úrovni buněk

NŠ, KH: PřF UPOL, KMAAM, 17. listopadu 12, 771 46 Olomouc
AA: Erasmus Universiteit Rotterdam, The Netherlands
JPA: Biomathematics and Statistics Scotland, Edinburgh, UK
PF: Vienna University of Technology, Austria

Nikola.Stefelova@seznam.cz

Mnohorozměrná data bývají obvykle uspořádána do matice s pozorováními v řádcích a proměnnými ve sloupcích. Obyčejné robustní odhady jsou navrženy tak, aby byly schopny vypořádat ses případy, kdy máme několik pozorování, které se (jako celek) odchylují od většiny. Tento přístup však může vést k významné ztrátě informace v situacích, kdy se odlehlost projevuje na úrovni jednotlivých buněk a ovlivňuje velkou část pozorování. Další problém nastává, pokud pracujeme s daty kompozičního charakteru. V tom případě je veškerá relativní informace pro statistickou analýzu obsažena v poměrech mezi složkami kompozice. Pak se např. špatně naměřená hodnota jedné složky projeví ve vztahu k ostatním složkám kompozice a dochází ke zkreslení výsledků. Cílem tohoto příspěvku je představit metodu robustní kompoziční regrese, která by byla schopna efektně vyřešit situaci s oběma možnostmi odlehlosti (tj. na celkové i buněčné úrovni). Ve stručnosti, odlehlé prvky matice jsou nejprve odfiltrovány a nahrazeny vhodnými hodnotami. Poté se pro nový datový soubor provede kompoziční MM-regrese. Z důvodů nejistoty, ke které dochází při imputování hodnot, se použije vícenásobná imputace. Výkonnost navrženého postupu je demonstrována na simulovaných i reálných biologických datech.

Zdeněk Šulc

nomclust 2.0: Balíček pro shlukování objektů charakterizovaných kategoriálními proměnnými

VŠE v Praze, náměstí Winstona Churchilla 1938/4, 130 67 Praha 3

zdenek.sulc@vse.cz

Balíček *nomclust* pro systém R kompletně pokrývá hierarchické shlukování kategoriálních datových souborů od výpočtu matice nepodobností pomocí jedné z 13 měr nepodobnosti po ohodnocení výsledných shluků prostřednictvím šesti hodnotících kritérií. Tento příspěvek představuje novou verzi tohoto balíčku, která obsahuje dvě výrazná vylepšení. První je věnováno optimalizaci výpočtu matice nepodobností, u kterého časová náročnost roste se čtvercem počtu pozorování. V případě kategoriálních souborů jsou nepodobnosti mezi dvěma objekty charakterizovanými určitou kombinací kategorií často počítány opakovaně. Příspěvek představuje algoritmus vedoucí k redukci rozměru této úlohy, a tím k výrazně kratšímu výpočetnímu času. Druhým vylepšením je přidání pěti hodnotících kritérií (modifikace AIC a BIC pro kategoriální data založené na entropii a mutabilitě, BK index), která slouží pro identifikaci optimálního počtu shluků. Na základě analýzy provedené na generovaných souborech se známým počtem shluků poskytoval BK index velmi dobré výsledky, a to jak mezi novými, tak mezi původními hodnotícími kritérii.

An R package *nomclust* completely covers hierarchical clustering of categorical datasets from a computation of a dissimilarity matrix using one of 13 dissimilarity measures to final cluster evaluation using six evaluation criteria. This contribution presents a new version of this package that contains two substantial enhancements. The first one deals with the optimization of a dissimilarity matrix determination which is a computationally demanding process, where the calculation time increases with the squared number of objects clustered. In categorical datasets, dissimilarities between two objects characterized by the certain combination of categories are often computed repeatedly. Thus, the contribution presents an algorithm that reduces the dimension of this task which leads to considerably shorter computational time. The second enhancement is adding five evaluation criteria (modifications of AIC and BIC for categorical data based on entropy and mutability, BK index), which serve for the optimal number of clusters identification. Based on the analysis performed on generated data with the known number of clusters, the BK index performed very well both among the new and original evaluation criteria.

Marie Turčičová, Jan Mandel, Kryštof Eben

Modelování kovariancí pro EnKF

MFF UK, KPMS, Sokolovská 83, 186 75 Praha 8
Ústav informatiky AV ČR, Pod Vodárenskou věží 271/2, 182 07 Praha 8

turcic@karlin.mff.cuni.cz

Ensemblový Kalmanův filtr (EnKF) je standardní metodou asimilace dat ve vysoké dimenzi. Tento filtr představuje Monte Carlo aproximaci klasického Kalmanova filtru, kdy je kovarianční matice nahrazena výběrovou kovarianční maticí vypočítanou z ensemblu členů, o nichž se často předpokládá, že tvoří náhodný výběr. Počet členů ensemblu je obvykle velmi malý v porovnání s jejich dimenzí a výběrová kovarianční matice má proto malou hodnost, navíc obsahuje tzv. rušivé korelace (spurious correlations). Použitím vhodného nízkoparametrického modelu pro kovarianční matici lze dosáhnout efektu regularizace a tím vylepšit vlastnosti kovariance vstupující do dalšího kroku filtru. V příspěvku budou prezentovány kovarianční modely tohoto typu, které vycházejí z typických vlastností kovariančních matic pro meteorologická data. Rovněž bude diskutován vliv kovariančního modelu na chybu EnKF.

Ondřej Vencálek

Klasifikace pomocí hloubky dat – nové nápady

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

ondrej.vencalek@upol.cz

V příspěvku se budeme věnovat možnostem použití hloubky dat při řešení úlohy klasifikace. V posledních dvou desetiletích bylo navrženo mnoho klasifikátorů využívajících hloubku dat. Pouze u některých z nich však byla dokázána jejich optimalita ve smyslu nejmenší možné celkové pravděpodobnosti chybného zařazení (tzv. Bayesovská optimalita). My se zamyslíme nad hodnotícími kritérii úspěšnosti klasifikace. Ne vždy totiž musíme trvat na co možná nejmenším celkovém počtu chyb, neboť různé chyby mohou být různě nákladné. Navrhujeme založit ztrátovou funkci (vyjadřující cenu za chybnou klasifikaci) na základě hloubky dat, tedy považovat chybné zařazení bodů v blízkosti centra distribuce za více závažné než chybné zařazení odlehlých pozorování. Studujeme tvar a vlastnosti klasifikátoru založeného na této myšlence.

Petr Volf

Využití směsí distribucí v modelování doby do poruchy

ÚTIA AV ČR, Praha 8

volf@utia.cas.cz

Rozdělení získaná jako směs jiných a jednodušších distribucí jsou používaná v mnoha oblastech, v případech, kdy jednoduché rozdělení nepopisuje data dostatečně. Směsi jsou vlastně také v pozadí některých shlukovacích metod. Zde tedy je distribuční funkce směsi konvexní kombinací distribučních funkcí komponent. Nejpopulárnější je zřejmě směs několika normálních rozdělení. V analýze spolehlivosti, konkrétně v modelování doby do poruchy, je jednou ze základních charakteristik riziková funkce (či intenzita poruch). Hraničním modelem je exponenciální rozdělení, které má intenzitu konstantní. Velice často má intenzita během celého životního cyklu (součástky či zařízení) tzv. vanovitý tvar, tj. zpočátku klesá, pak je vpodstatě konstantní, ke konci roste. Není proto divu, že byly navrženy i modely, jejichž riziková funkce je kombinací (tedy směsí) několika monotónních funkcí. Otázka tedy je, který typ směsi je pro nějaký konkrétní případ vhodnější. Tímto problémem se budu zabývat, tj. jednak odhadováním parametrů směsí v obou případech, jednak porovnáváním, pomocí věrohodnosti či jiných kritérií (založených např. na vzdálenostech Kolmogorova-Smirnova a podobných), který model datům víc odpovídá. Další možností, která se pro směsi rizikových funkcí nabízí, je kombinovat jejich "míchání"s detekcí bodů změny, tj. bodů, kde se mění charakter rizikové funkce.

Viktor Witkovský

A Note on computing the exact distribution of selected multivariate test criteria

Institute of Measurement Science, Slovak Academy of Sciences, Dúbravská cesta 9, 841 04 Bratislava

witkovsky@savba.sk

Application of the exact statistical inference frequently leads to a non-standard probability distributions of the considered estimators or test statistics. The exact distributions of many estimators and test statistics can be specified by their characteristic functions. The characteristic function represents complete characterization of the distribution of the random variable. However, analytical inversion of the characteristic function, if possible, frequently leads to a complicated and computationally rather strange expressions for the corresponding distribution function (CDF/PDF) and the required quantiles.

Here we advocate to use the method based on numerical inversion of the characteristic functions, as implemented, e.g., in the MATLAB toolbox `CharFunTool` (The Characteristic Functions Toolbox), available at the web page `https://github.com/witkovsky/CharFunTool/`. The applicability of the approach is illustrated by computing the exact null and non-null distributions of selected multivariate test criteria. In particular, we shall discuss the distribution of the Bartlett's test criterion for testing homogeneity of variances in several normal populations and the Wilks's Λ-distribution used in testing hypotheses in multivariate statistical analysis.

Poďakovanie: Supported by the Slovak Research and Development Agency, project APVV-15-0295.

Gejza Wimmer a Petra Ráboňová

Konfidenčné oblasti pre koeficienty kalibračnej funkcie

P.R.: Masarykova Univerzita, Ústav matematiky a statistiky, Kotlářská 2, Brno a
G.W.: Univerzita Mateja Bela, Fakulta prírodných vied, Tajovského 40, Banská Bystrica; Matematický ústav SAV, Štefánikova 49, Bratislava

324037@mail.muni.cz, wimmer@mat.savba.sk

Príspevok pojednáva o modeli komparatívnej parametrickej kalibrácie, t.j. analyzuje situáciu, keď veličiny merania kalibrovaným aj kalibračným prístrojom (etalónom) sú zaťažené náhodnými chybami merania. Kalibračná funkcia je polynóm. Tento model je modelom s chybami v premenných (errors-in-variables model). Po linearizovaní môžeme kalibračný model považovať za (lineárny) model nepriamych meraní s podmienkou II.typu na parametre 1. rádu. Pri neznalosti variančných koeficientov je možné dvomi spôsobmi použiť aproximácie typu Kenwarda a Rogera na získanie konfidenčných oblastí pre koeficienty kalibračnej funkcie. Na malej simulačnej štúdii porovnávame tieto dve konfidenčné oblasti.

Poďakovanie: Práca bola podporená grantom APVV-15-0295.

Marta Žambochová

Algoritmy pro shlukování prostorových dat

FSE UJEP, KMI, Moskevská 54, 400 96 Ústí n. L.

marta.zambochova@ujep.cz

Velmi častým typem dat jsou tzv. prostorová data. Prostorová data jsou charakteristická tím, že obsahují jednak prostorovou složku, která popisuje polohu a tvar jednotlivých objektů, a jednak složku s informacemi o vlastnostech objektů. S těmito daty se setkáváme např. v geografických informačních systémech (GIS), biomedicínských oblastech či v zemědělských vědách. K analýze těchto dat se bohužel nehodí všechny klasické metody statistického zpracování. Mezi nejzávažnější bariéry pro použití některých klasických metod je například prostorová autokorelace, nerespektování prostorových vztahů, ale i veliký rozsah těchto dat. Příspěvek přináší popis a srovnání vybraných algoritmů shlukové analýzy vhodných pro práci s prostorovými daty.