MARKEROVÁ STATISTIKA

Jiří Knížek

Leden 2013

Statistické algoritmy, zaměřující se na závislostní identifikaci (bio)markerů a určování diagnózy

Knizek, J. 2011. *Marker Statistics I.: Regression analysis of dependences in medicine and molecular biology*, VDM Publishing House Ltd., Mauritius.

Testy v soustavě vícenásobných lineárních regresí

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_1 & & & \\ & \boldsymbol{X}_2 & & \\ & & \ddots & \\ & & & \boldsymbol{X}_M \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_M \end{bmatrix} + \begin{bmatrix} \boldsymbol{e}_1 \\ \boldsymbol{e}_2 \\ \vdots \\ \boldsymbol{e}_M \end{bmatrix}$$

$$(\text{or briefly } \; \boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{e})$$

$$H_0 : \; \boldsymbol{R}\beta = \mathbf{r}$$

Lze testovat libovolné vzájemné lineární vztahy mezi jednotlivými subvektory parametrů $\beta_1, \beta_2, \ldots, \beta_M$

Testy v soustavě vícenásobných lineárních regresí

$$\lambda_F = \frac{(\mathbf{r} - \boldsymbol{R}\hat{\boldsymbol{\beta}})'(\boldsymbol{RCR}')^{-1}(\mathbf{r} - \boldsymbol{R}\hat{\boldsymbol{\beta}}) / J}{(\mathbf{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I})(\mathbf{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) / (MT - K)} \sim F_{(J,\, MT-K)},$$

$$\boldsymbol{C} = [\boldsymbol{X}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I})\boldsymbol{X}]^{-1},$$

Judge, G. G, Griffiths, W. E, Hill, R. C, Lutkepohl, H., Tsoung-Chao, L., 1985, *The Theory and Practice of Econometrics,* J. Wiley, New York.

Gatignon, H., 2003, *Statistical Analysis of Management Data.* Kluwer Academic Publishers (New York, Boston, Dordrecht, London, Moscow).

Testy v soustavě ortogonálních polynomických regresí = testy v soustavě spektrálních průběhů

$$H_0 : \boldsymbol{k}\boldsymbol{\eta}(x) = \boldsymbol{r}(x)$$

definiční matice

$$\boldsymbol{k} = \begin{pmatrix} k_{1,1} & k_{1,2} & \dots & k_{1,M} \\ k_{2,1} & k_{2,2} & \dots & k_{2,M} \\ \vdots & \vdots & \ddots & \vdots \\ k_{J,1} & k_{J,2} & \dots & k_{J,M} \end{pmatrix}$$

vektor spektrálních průběhů v abscise x

$$\boldsymbol{\eta}(x) = (\eta_1(x), \eta_2(x), \dots, \eta_M(x))'$$

Testy v soustavě ortogonálních polynomických regresí = testy v soustavě spektrálních průběhů

$$p(x) = 1 - \mathrm{fcdf}(\hat{\lambda}(x),\ J, M') \qquad M' = MT - M(\Re + 1)$$

$$power\ of\ test(x) = 1 - \mathrm{ncfcdf}(\mathrm{finv}(1\ -\ \alpha, J, M'),\ J, M', \hat{\delta}(x, \hat{\boldsymbol{\beta}}))$$

$$\hat{\lambda}(x) = \frac{\hat{\delta}(x, \hat{\boldsymbol{\beta}}) \Big/ J}{s_{\boldsymbol{e}}^2}$$

$$\hat{\delta}(x, \hat{\boldsymbol{\beta}}) = (\boldsymbol{r}(x) - \boldsymbol{k}\,\hat{\boldsymbol{\eta}}(x))' \left\{ [\boldsymbol{k}\boldsymbol{X}(x)]\, \hat{\boldsymbol{B}}\, [\boldsymbol{k}\boldsymbol{X}(x)]' \right\}^{-1} (\boldsymbol{r}(x) - \boldsymbol{k}\,\hat{\boldsymbol{\eta}}(x))$$

$$s_{\boldsymbol{e}}^2 = (\boldsymbol{y} - \boldsymbol{\Psi}\,\hat{\boldsymbol{\beta}})' (\hat{\boldsymbol{\Sigma}}^{-1} \otimes \boldsymbol{I})\, (\boldsymbol{y} - \boldsymbol{\Psi}\,\hat{\boldsymbol{\beta}}) \Big/ M'$$

Testy v soustavě ortogonálních polynomických regresí = testy v soustavě spektrálních průběhů

Optimalizace stupňů (ortogonálních) spektrálních polynomů

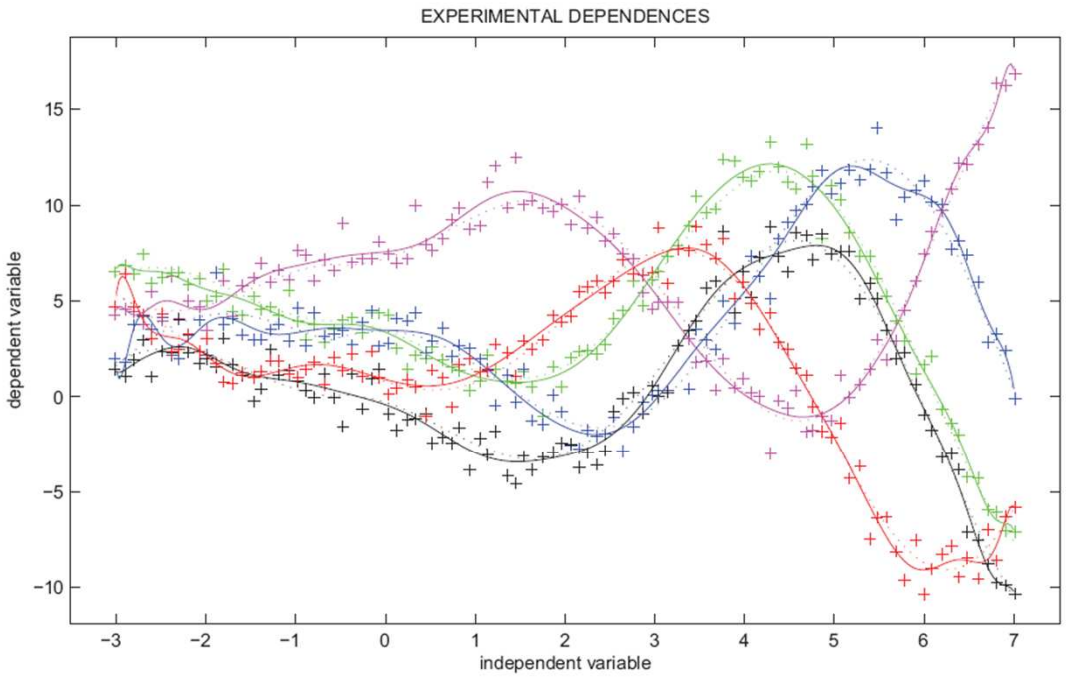


Figure 1. Demonstration of effectiveness of the process of searching for *estimation of regression orthogonal polynomials* on simulated data. Dotted lines - simulated courses of *functional dependences*; plus signs – randomized (3%) courses of *functional dependences*; solid lines - statistical estimation of courses of *functional dependences* based on randomized courses of *functional dependences*.

Testy v soustavě ortogonálních polynomických regresí = testy v soustavě spektrálních průběhů

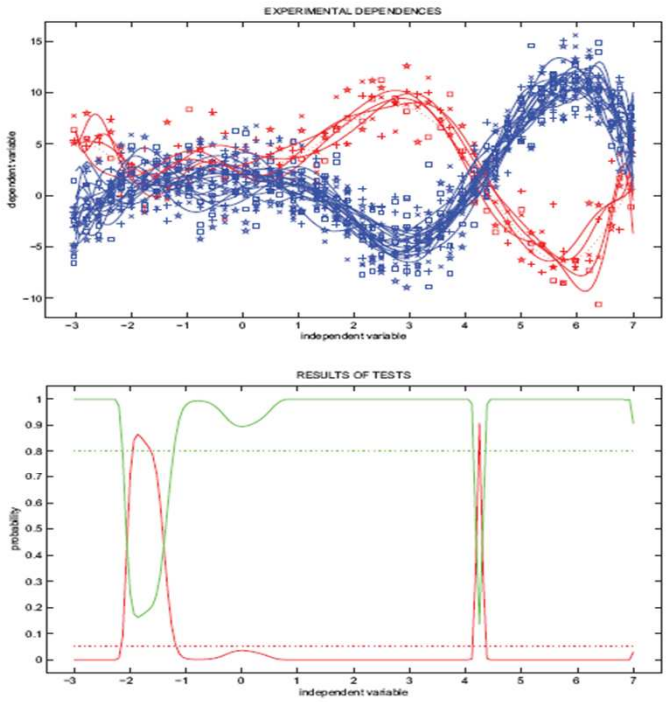


Figure 3. (3a-upper and 3b-lower) Demonstration of efficiency of *the extended paired dependences' test*. Upper graph: two simulated groups (red (e.g. diseased) and blue (e.g. healthy)) of stochastic *functional dependences*. Dotted lines - simulated "ideal" courses of *functional dependences*; plus signs, squares, crosses and five-pointed stars – randomized (25%) discrete courses of *functional dependences*; solid lines - statistical estimation of courses of *functional dependences* based on randomized courses of *functional dependences*. Lower graph: Corresponding (red solid line) dependence of *p-value* $p(x)$ on the appropriate independent variable x. Requested test *significance level* α is a red dotted line. Corresponding (green solid line) dependence of $power\ of\ test(x)$ on the appropriate independent variable x. Requested $power\ of\ test(x)$ is a green dotted line (Cohen, 1988; Daly and Bourke, 2000).

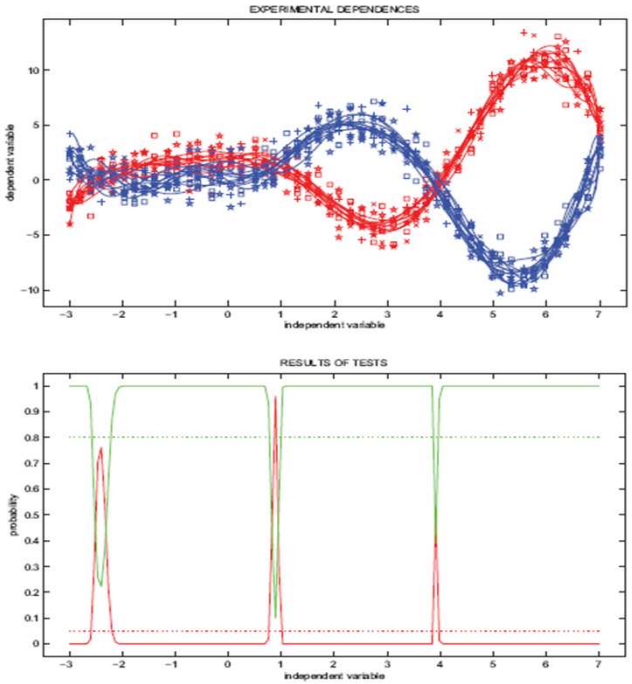


Figure 2. (2a-upper and 2b-lower) Demonstration of efficiency of *the extended paired dependences' test*. Upper graph: two simulated groups (red (e.g. diseased) and blue (e.g. healthy)) of stochastic *functional dependences*. Dotted lines - simulated "ideal" courses of *functional dependences*; plus signs, squares, crosses and five-pointed stars – randomized (25%) discrete courses of *functional dependences*; solid lines - statistical estimation of courses of *functional dependences* based on randomized courses of *functional dependences*. Lower graph: Corresponding (red solid line) dependence of *p-value* $p(x)$ on the appropriate independent variable x. Requested test *significance level* α is a red dotted line. Corresponding (green solid line) dependence of $power\ of\ test(x)$ on the appropriate independent variable x. Requested $power\ of\ test(x)$ is a green dotted line (Cohen, 1988; Daly and Bourke, 2000).

Testy v soustavě ortogonálních polynomických regresí = testy v soustavě spektrálních průběhů

Knizek, J., Sindelar, J., Beranek, L., Vojtesek, B., Nenutil, R., Brozkova, K., Drazan, V., Hubalek, M. and Kubacek, L. 2008a. Power function for tests of null hypotheses on mutual linear regression functions, *Bulletin of Statistics and Economics* **2**(S08): 26-33.

Knizek, J., Sindelar, J., Pulpan, Z., Vojtesek, B., Nenutil, R., Brozkova, K., Drazan, V., Hubalek, M. and Beranek, L. 2008b. Test of the hypothesis that one group of dependences is consistent with another group of dependences, *Bulletin of Statistics and Economics* **2**(A08): 2-18.

Knížek·J.,·Beránek·L.,·Bouchal·P.,·Vojtěšek·B.,·Nenutil·R.,·and·Tomšík·P.·*Extended·Paired·Dependences'·Test*,·International·Journal·of·Applied·Mathematics·and·Statistics;·Vol.·37;·Issue·No.·7;·Year·2013,·ISSN·0973-1377·(Print),·ISSN·0973-7545·(Online);·Copyright·©·2013·by·CESER·Publications¶

Identifikace biomarkerů simultánními testy

$$H_0 : \boldsymbol{k}\boldsymbol{\eta}(x) = \boldsymbol{r}(x) \qquad \boldsymbol{k} = \begin{pmatrix} k_{1,1} & k_{1,2} & \dots & k_{1,M} \\ k_{2,1} & k_{2,2} & \dots & k_{2,M} \\ \vdots & \vdots & \ddots & \vdots \\ k_{J,1} & k_{J,2} & \dots & k_{J,M} \end{pmatrix}$$

Konstrukce nulové hypotézy H_0 je postavena tak (Judge et al., 1985), že se (případně) zamítá ve prospěch H_A již tehdy , není-li splněna jediná z J rovností $k\eta(x) = r(x)$. Z biofyzikálních důvodů, však, potřebujeme, aby děj platil kompletně. Tj., potřebujeme zamítat H_0 ve prospěch H_A takové, že z J rovností $k\eta(x) = r(x)$ neplatí všechny zároveň. Východiskem je test přísl. simultánních hypotéz:

Identifikace biomarkerů simultánními testy

$$H_0^j:\ \boldsymbol{k}^j_{(1\times M)}\boldsymbol{\eta}_{(M\times 1)}(x)=\boldsymbol{r}^j_{(1\times 1)}(x)=r^j(x)$$

$$j=1,2,\ldots,J$$

$$p_j(x)<\alpha/J,\qquad j=1,2,\ldots,J$$

$$1-\beta_j(x)\geq convention\ limit,\qquad j=1,2,\ldots,J$$

$\boldsymbol{k}^j_{(1\times M)}$ ***je příslušný řádek definiční matice*** $\boldsymbol{k}$

Identifikace biomarkerů simultánními testy

Pro danou abscisu x, se pak výsledná („nejnepříznivější“) p-hodnota a příslušná síla testu spočítají pomocí vztahů:

$$p_{\text{result}}(x) = p_{j'}(x) = \max_{j=1,2,\ldots,J} p_j(x)$$

$$power\ of\ test_{\text{result}}(x) = power\ of\ test_{j'}(x)$$

Knizek, J., Sindelar, J., Vojtesek, B., Bouchal, P., Nenutil, R. and Beranek, L. 2010. Identification of markers by simultaneous tests in a set of quantifying dependences, *International Journal of Statistics and Economics* **5**(A10): 12–20.

Numericko-matematické aspekty použitých algoritmů

Klíčovou ingrediencí statistických závislostních algoritmů je používání numericky stabilního způsobu generování hodnot normalizovaných ortogonálních polynomů na diskrétní množině bodů. Od této algoritmické vlastnosti se pak odvíjí kvalita/nekvalita polynomiální aproximace naměřených tabelárních funkcí. My používáme velmi účinný Arnoldiho algoritmus s reortogonalizací. Tento algoritmus produkuje polynomy s malou nepřesností, srovnatelnou se strojovou nepřesností daného programovacího systému (matlab).

Knizek, J., Tichy, P., Beranek, L., Sindelar, J., Vojtesek, B., Bouchal, P., Nenutil, R. and Dedik, O. 2010. Note on generating orthogonal polynomials and their application in solving complicated polynomial regression tasks, *International Journal of Mathematics and Computation* **7**(J10): 48–60.

Numericko-matematické aspekty použitých algoritmů

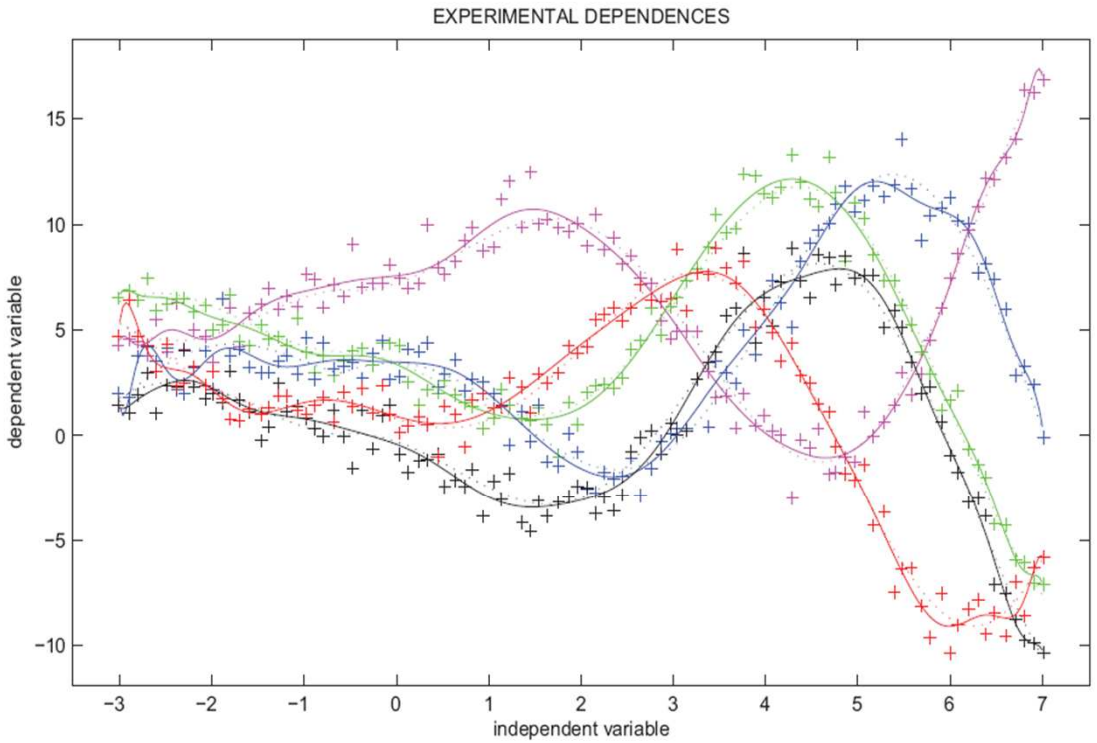


Figure 1. Demonstration of effectiveness of the process of searching for *estimation of regression orthogonal polynomials* on simulated data. Dotted lines - simulated courses of *functional dependences*; plus signs – randomized (3%) courses of *functional dependences*; solid lines - statistical estimation of courses of *functional dependences* based on randomized courses of *functional dependences*.

Markerové určování diagnózy

Náš formální matematický přístup (testy v soustavě spektrálních průběhů s tzv. definiční maticí ***k****) přímo nabízí sestrojení programového systému, který by umožňoval pomocí přiměřeně velkého souboru „historických“ dat a dat „aktuálních“ (např. od daného pacienta) určovat diagnózu.*

Jsou zde dva základní přístupy rozhodování:

1) *Testování „hlavních simultánních nulových hypotéz“:*

$$H_0^{\text{main},\,j}\,,\quad j = 1,2,\ldots,J\,,$$

2) *Testování „doplňkových simultánních nulových hypotéz“:*

$$H_0^{\text{complementary},\,j}\,,\quad j = 1,2,\ldots,J\,,$$

Markerové určování diagnózy

1) V případě testování „hlavních simultánních nulových hypotéz“ pacient vystupuje v „roli zdravého“ a v případě zamítnutí nulových hypotéz $H_0^{\text{main},\,j}$, $j = 1,2,\ldots,J$, *pacient pravděpodobně není zdráv a tedy by měl být léčen.*

2) V případě testování „doplňkových simultánních nulových hypotéz“ pacient vystupuje v „roli nemocného“ a v případě nezamítnutí nulových hypotéz $H_0^{\text{complementary},\,j}$, $j = 1,2,\ldots,J$, *je zde zpráva o tom, že pacient pravděpodobně není zdráv, a tedy by měl být léčen.*

Markerové určování diagnózy

Formálně matematicky:

Schematically, we can express the testing of the *main simultaneous null hypotheses* of the type

$$\left.\begin{aligned} & H_0^{\text{main}} : \overbrace{\Delta(\text{diseased} - \text{healthy})}^{\text{formerly}} = \overbrace{\Delta(\text{diseased} - \text{patient})}^{\text{topically}} \\ & \text{or} \\ & H_0^{\text{main}} : \underbrace{\Delta(\text{diseased} - \text{healthy})}_{\text{formerly}} - \underbrace{\Delta(\text{diseased} - \text{patient})}_{\text{topically}} = 0. \end{aligned}\right\} \quad (1.2)$$

It means that we compare (we test) any formerly measured "diseased - healhty" difference with any topically measured "diseased - patient" difference, where a patient performs as if *in the role of a healthy man.*

Quite analogically, we can express schematically that we test *the complementary simultaneous null hypotheses* of type

$$\left.\begin{aligned} & H_0^{\text{complementary}} : \overbrace{\Delta(\text{diseased} - \text{healthy})}^{\text{formerly}} = \overbrace{\Delta(\text{patient} - \text{healthy})}^{\text{topically}} \\ & \text{or} \\ & H_0^{\text{complementary}} : \underbrace{\Delta(\text{diseased} - \text{healthy})}_{\text{formerly}} - \underbrace{\Delta(\text{patient} - \text{healthy})}_{\text{topically}} = 0. \end{aligned}\right\} \quad (1.3)$$

It means that we compare (we test) any formerly measured "diseased - healhty" difference with any topically measured "patient - healthy" difference, where a patient performs as if *in the role of a diseased man.*

Markerové určování diagnózy

Knizek, J., Sindelar, J., Vojtesek, B., Bouchal, P., Nenutil, R., Beranek, L. and Dedik, O. 2010. Using markers to aid decision making in diagnostics, *International Journal of Tomography and Statistics* **16**(W11): 41–55.

„Odlehlé chování“ některých spektrálních průběhů (pacientů) = obrovská algoritmická komplikace

Je známo, že některé biomarkery „nefungují“ u populace zcela 100%ně. Tj., některé biomarkery se chovají jako biomarkery jen u určité části (např. 70%) populace apod. Tj., některé MS-spektrální průběhy se „chovají odlehle“.

Vyvstává zde otázka, jak respektovat tento problém algoritmicky?

Na první pohled se řešení může jevit snadným. Přece, postupně, krok za krokem, uvažovat všechny myslitelné kombinace spektrálních průběhů a to tak, že je jsou jednotlivé spektrální průběhy (nebo jejich skupiny) systematicky postupně vynechávány. Každá takováto sestava je pak vyhodnocována za účelem identifikace biomarkerů. Moderní super-výkonné počítače by to měly zvládnout!

„Odlehlé chování“ některých spektrálních průběhů (pacientů) = obrovská algoritmická komplikace

Lze dokázat, že se vzrůstajícími základními parametry sytému† počty takovýchto sestav rychle vzrůstají k astronomickým hodnotám a tudíž tento přístup je použitelný jen ve velmi omezené míře:

	$q_{\min}$					
	100%	90%	80%	70%	60%	50%
$k = 10$	1	11	56	176	386	638
$k = 20$	1	211	6196	60460	263950	616666
$k = 30$	1	4526	768212	22964087	194129627	614429672
$k = 40$	1	102091	100146724	9119901052	147437500478	618679078298
$k = 50$	1	2369936	13432735556	3715721875476	114075475473136	626155256640188

Table 2. Sums $S_{k,q_{\min}}$ of all conceivable indexes $j_1, j_2, \ldots, j_m = 1, 2, \ldots, k$, $(k \leq m)$.

Knizek, J., Sindelar, J., Vojtesek, B., Bouchal, P., Nenutil, R. and Beranek, L. 2012. The Marker Statistic's Problem with Occurrence of Tested Phenomena's Outlying Behavior, *International Journal of Statistics and Economics* **9**(A12): 83-89.

† ***počet spekter, počet vynechaných členů atd.***

Gnostické řešení případů „odlehlého chování“

Protože (současná) robustní statistika neposkytuje vhodné regresní modely, které by umožňovaly provádění testů hypotéz v soustavě spektrálních průběhů s případy „odlehlého chování“, uchýlili jsme se ke gnostickému řešení tohoto problému.

Gnostická Teorie Dat (GTD) je alternativou statistiky (Kovanic 1986). GTD je určena pro odvozování algoritmů na zpracování dat za praktických okolností, kdy není dostatek dat, kdy data jsou kontaminována silnými neurčitostmi a kdy matematicko-statistický model dat a jejich neurčitostí není znám anebo neexistuje. GTD není založena na statistických předpokladech.

Gnostické řešení případů „odlehlého chování“

GTD formuluje matematický model jednotlivých datových neurčitostí na základě jednoduchého metrologického axiomu. Teorie malých datových souborů pak vyplývá z teorie jednotlivých dat a z kompozičního zákona, určujícího způsob skládání neurčitostí jednotlivých dat. GTD produkuje významně robustní algoritmy pro zpracování dat.

Kovanic, P. 1986. A New Theoretical and Algorithmical Basis for Estimation, Identification and Control. *Automatica* **22**(6): 657-674.

Knizek J, Beranek L, Bouchal P, Vojtesek B, Nenutil R, Tomsik P, 2013, *Gnostic Solution of the Marker Statistics' Problem with Occurrence of Tested Phenomena's Outlying Behavior*, International Journal of Ecological Economics & Statistics, Volume 29, Number 2; ISSN 0973-7537.

Primární třídící znak pro spektra v oblasti (potencionálního) daného (bio)markeru je píkovitý tvar alespoň jednoho z dané sady spektrálních průběhů

(Složená) nulová hypotéza H_0 *o koeficientu determinace:*

$$H_0 : R^2 \leq R_0 \,, \quad H_A : R^2 > R_0 \,,$$

(např.) pro funkci

$$y = c + \kappa \, exp[-(x - \mu)^2 / \delta] \,.$$

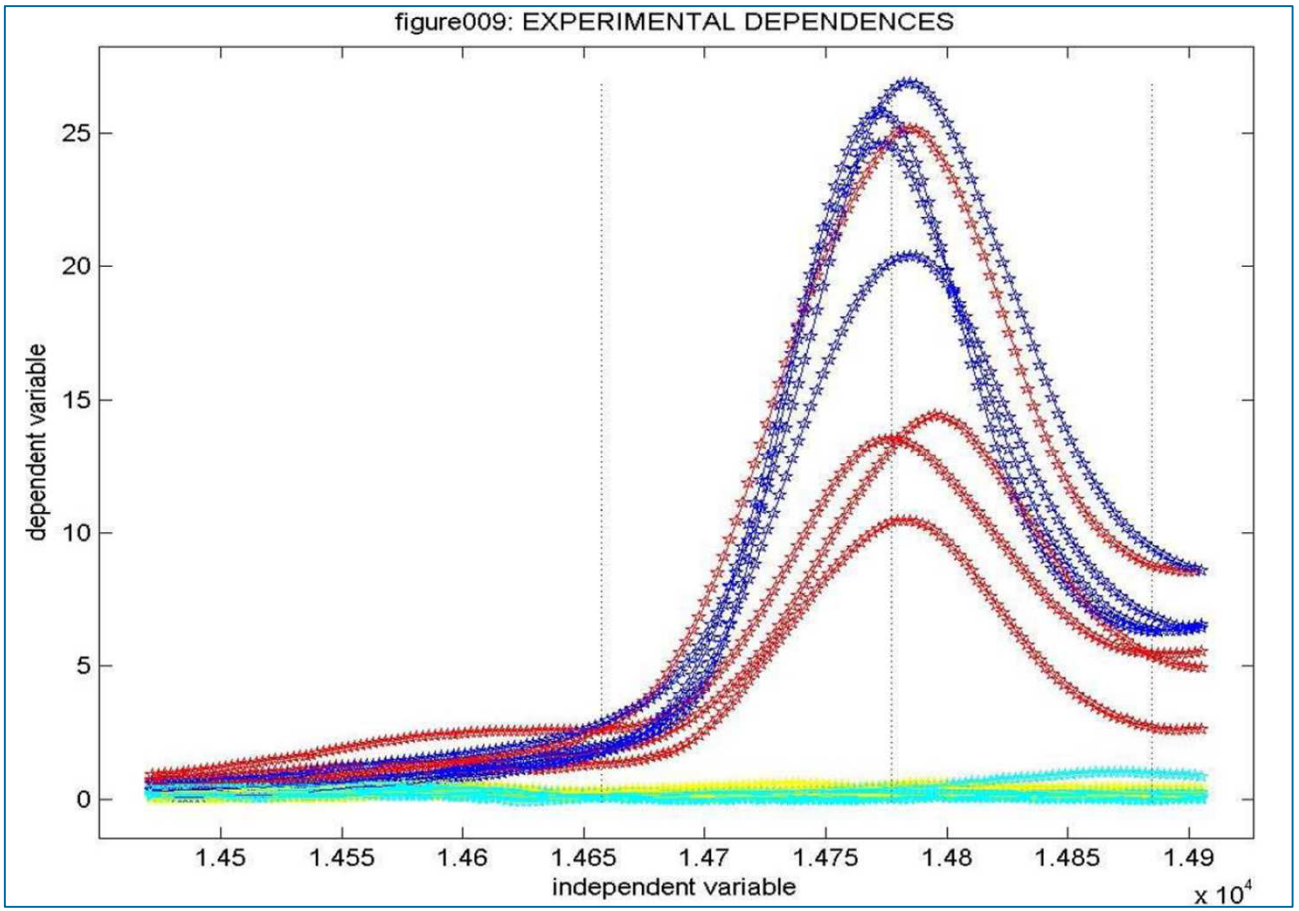
Druhotný třídící znak pro spektra v oblasti (potencionálního) daného (bio)markeru jsou pak pravidla, vyplývající z (bio)chemické či (bio)fyzikální logiky daného systému spektrálních průběhů

V případě, že nebyla zamítnuta (složená) nulová hypotéza H_0 *o koeficientu determinace:*

$$H_0 : R^2 \leq R_0 \text{ ,}$$

uvažujeme i případ, kdy část populace je inaktivní vůči danému biomarkeru

Druhotný třídící znak pro spektra v oblasti (potencionálního) daného (bio)markeru jsou pak pravidla, vyplývající z (bio)chemické či (bio)fyzikální logiky daného systému spektrálních průběhů



Pro primární hledání píkovitého tvaru v spektrálních průbězích byl vyvinut speciální algoritmus „PEAK“ založený na systematickém zjišťování sekvencí:

pro vzestupnou část píku:

```
12  % output scalars:
13  % r                    ... result of ascertainment
14  %                          =1 when:
15  %                          y(1) < y(2) < y(3) < ... < y(n-1) < y(n)
16  %                          else: =0
```

pro sestupnou část píku:

```
12  % output scalars:
13  % r                    ... result of ascertainment
14  %                          =1 when:
15  %                          y(1) > y(2) > y(3) > ... > y(n-1) > y(n)
16  %                          else: =0
17  %
```

Gnostické řešení případů „odlehlého chování“

Teprve v rámci autonomních skupin spekter, vzniklých pomocí primárního a druhotného třídění, je možné použít gnostickou shlukovou analýzu, a sice, na dvou úrovních:

1. produkuje tzv. „tvrdé shluky“ (opt. par. měř.)
2. produkuje tzv. „měkké shluky“ (syst. sniž. par. měř.)

v x-ové souřadnici maxima daného (potencionálního) (bio)markeru. Pomocí uvedených prostředků je pak možné si představit určování různých stupňů virulence nemoci.

Pořadí úkonů pro identifikaci potenciálních (bio)markerových oblastí v sadě (např. SELDI-TOF hmotnostních) spekter:

1) *Hledání píkovitého tvaru v spektrálních průbězích (algoritmus „PEAK“)*
2) *Druhotný třídící znak pro spektra v oblasti (potencionálního) daného (bio)markeru – (bio)markerové chování/nechování; testy hypotéz o koeficientu determinace*
3) *Identifikace biomarkerů simultánními testy – 1. stupeň*
4) *Gnostické řešení případů „odlehlého chování“ („tvrdé clustery“)*
5) *Identifikace biomarkerů simultánními testy – 2. stupeň*
6) *Gnostické řešení případů „odlehlého chování“ („měkké clustery“)*
7) *Identifikace biomarkerů simultánními testy – 3. stupeň*

V historii velmi pravděpodobně došlo mnohokráte k případu, že z velmi nákladných dat nebylo zdaleka „vytěženo“ možné maximum informace v důsledku toho, že jejich zpracování nebylo provedeno adekvátně důsledným matematicko-statistickým algoritmem.

Knizek 2012 __l.pdf

Matematika zpětně ovlivňuje plán experimentování

Speciální konstrukce testu: „Rozšířený Párový Test Závislostí“. K dispozici jsou experimentální vzorky (závislosti) v počtu $M_{j,\text{diseased}}$ a zároveň experimentální vzorky v počtu $M_{j,\text{healthy}}$, kde $j = 1,2....,J$, pro j-té individuum (osoba, pacient, laboratorní zvíře, mikroorganizmus atd.). Obecně počty $M_{j,\text{diseased}}$ a $M_{j,\text{healthy}}$ nemusí být vzájemně totožné. Přitom, symbol $M_{j,\text{diseased}}$ reprezentuje počet vzorků získaných např. z nádorové tkáně a symbol $M_{j,\text{healthy}}$ reprezentuje počet vzorků získaných ze zdravé tkáně u téže osoby. Symbol J reprezentuje počet testovaných individuí (osoby, pacienti, laboratorní zvířata, mikroorganizmy atd.). Tato testová konstrukce je relativně snadno dostupná přístupem pomoci tzv. definiční matice. V praxi, se potřeba tohoto testu může vyskytovat velmi často z nejrůznějších důvodů. Např. některé vzorky se nepodařilo dokončit z experimentálních (anatomických, biochemických apod.) důvodů. Jindy, získání dalších vzorů může být finančně příliš nákladné nebo, prostě, neproveditelné (např. z důvodů etických).

Matematika zpětně ovlivňuje plán experimentování

Knížek J., Beránek L., Bouchal P., Vojtěšek B., Nenutil R., and Tomšík P. *Extended Paired Dependences' Test*, International Journal of Applied Mathematics and Statistics; Vol. 37; Issue No. 7; Year 2013, ISSN 0973-1377 (Print), ISSN 0973-7545 (Online); Copyright © 2013 by CESER Publications.

Výpočetní časy

***MetaCentrum:** 6124 procesorů v 393 počítačích (prosinec 2012), umožňuje paralelní výpočty jednotlivých segmentů*

***Data:** párové závislostní testy (10 pacientů): týden*

***Data:** nepárové závislostní testy (33 z. „nemocní“, 29 z. „zdraví“): asi 10 let*

***Algoritmus:** n. z. t. rozšířený o používání tzv. „řídkých matic“: týden*

Další nezávisle proměnná: čas τ

Test simultánních nulových hypotéz H_0^j , $j = 1,2,\ldots,J$,
o tom, že (pro danou abscisu x*) se vzájemně rovná skupina „nemocných“* $y_{i,\mathrm{nem}}(\tau)$ *a skupina „zdravých“* $y_{i,\mathrm{zdr}}(\tau)$ *spektrálních závislostí na čase* τ $(i = 1,2,\ldots,M)$
a zároveň jejich první a druhé derivace podle času: $\partial y_{i,\mathrm{nem}}(\tau)/\partial\tau$ *vs.* $\partial y_{i,\mathrm{zdr}}(\tau)/\partial\tau$ *a* $\partial^2 y_{i,\mathrm{nem}}(\tau)/\partial\tau^2$ *vs.* $\partial^2 y_{i,\mathrm{zdr}}(\tau)/\partial\tau^2$.

Další nezávisle proměnná: čas τ

Vychází z biofyzikálních předpokladů:
Čím se projevuje chování proteinů (daného autonomního biologického systému) v čase?
Samozřejmě tím, že protein ubývá a/nebo (zase) přibývá anebo jeho množství (resp. koncentrace) zůstává téměř konstantní. Jemněji je možné rozlišovat, zda rychlost ubývání nebo přibývání proteinu se zpomaluje (konkávní) nebo zrychluje (konvexní časový průběh koncentrace proteinu).

Další nezávisle proměnná: čas τ

Knizek, J. 2011. *Marker Statistics I.: Regression analysis of dependences in medicine and molecular biology*, VDM Publishing House Ltd., Mauritius.

Knížek J, Bergmann M, Šindelář J, Kovářová H (2004b). MIAPS - programový systém pro odhadování pořadí vzájemné podobnosti/nepodobnosti chování proteinů v čase. *Acta Medica (Hradec Králové) Supplementum*, 47(2), 131-135.