Andrej Pázman

Geometrické metody v matematickej štatistike

Pokroky matematiky, fyziky a astronomie, Vol. 33 (1988), No. 6, 314--326

Persistent URL: http://dml.cz/dmlcz/139273

Geometrické metódy v matematickej štatistike

Andrej Pázman, Bratislava

1. Úvod

Matematická štatistika, ako teória hromadného spracovania údajov a získavania informácie z týchto údajov, tradične vychádza z teórie pravdepodobnosti. Na údaje získané meraním, pozorovaním, prieskumom a pod. sa díva ako na určitý odraz skúmanej reality, ktorý je „zahmlený" radom podružných faktorov. Vplyv týchto faktorov považuje matematická štatistika za náhodný a v spleti údajov sa štatistik snaží nájsť určitý poriadok a vytiahnuť z nich užitočnú informáciu práve na základe poznatkov teórie pravdepodobnosti.

Základná všeobecná schéma, s ktorou pracuje matematická štatistika býva často nasledujúca. Stav pozorovaného resp. meraného objektu je charakterizovaný nejakým vektorom číselných parametrov $\theta = (\theta_1, \ldots, \theta_m)^T$. Pozorovateľ, resp. experimentátor získava empirickou, resp. experimentálnou cestou rôzne údaje odzrkadlujúce (nepriamo) stav objektu. Označme $y = (y_1, \ldots, y_N)^T$ tieto údaje usporiadané do vektora. Štatistik považuje y za realizáciu náhodného vektora, pričom vychádza z predpokladu, že pri známom vektore θ bolo by známe aj rozdelenie pravdepodobnosti $p(y \mid \theta)$ vektora y.

V skutočnosti však štatistik nepozná stav meraného objektu, t.j. nepozná θ, nanajvýš vie, že θ nemôže vybočiť z istej množiny Θ (nazývanej *parametrický priestor*). Úlohou štatistika je získať informáciu o skutočnej hodnote vektora θ, alebo o niektorých jeho komponentoch, a to práve na základe pozorovanej hodnoty vektora y. Jeho úloha je v istom zmysle opačná než úloha odborníka úzko špecializovaného na pravdepodobnosť. „Pravdepodobnostiar" by analyzoval rozdelenie náhodného vektora y pri danom θ, kdežto štatistik pátra po správnej hodnote θ na základe daného vektora y. Pritom pochopiteľne využíva to, čo pozná o rozdeleniach pravdepodobnosti vektora y pri rôznych *hypotetických* hodnotách $\theta \in \Theta$. Na rozdiel od „pravdepodobnostiara" štatistik teda pracuje s celou *triedou* hypotetických rozdelení pravdepodobností $\{p(y) \mid \theta)$: $\theta \in \Theta\}$.

Vychádzajúc zo znalostí takejto triedy, úlohou štatistika je vypracovať to, čomu hovoríme metódy štatistickej inferencie. Napríklad musí nájsť nejaké zobrazenie τ: $y \in Y \mapsto \tau(y) \in \Theta$, ktoré možno považovať v nejakom zmysle za dobrý odhad vektora parametrov θ (tu symbolom Y sme označili tzv. *výberový priestor*, t.j. množinu možných hodnôt vektora údajov y). V praxi to zvyčajne znamená, že štatistik vymyslí nejaký „vzorec", do ktorého dosadí napozorovaný vektor y a vypočítaný výsledok označí za hodnotu blízku skutočnej, ale neznámej hodnote θ. Pochopitelne, že v mnohých prípadoch pod slovom „vzorec" treba chápať aj algoritmus pre počítač alebo iný zložitý

postup. Stanovením odhadu vektora θ sa však práca štatistika nekončí: musí ešte charakterizovať presnosť takéhoto odhadu, v niektorých prípadoch musí hľadať vhodné miery množstva informácie získanej pozorovaním alebo musí vedieť predpovedať vlastnosti odhadov (ešte pred pozorovaním) a na základe toho navrhnúť optimálny experiment a pod. Súhrnom, práca štatistika (i teoretického štatistika) sa líši od práce odborníka úzko špecializovaného na teóriu pravdepodobnosti; nielen preto, že štastitik často pracuje s reálnymi údajmi, ale aj preto, že pracuje, ako už bolo povedené, s celou triedou hypotetických rozdelení pravdepodobnosti $\{p(y) \mid \theta)\colon \theta \in \Theta\}$.

Otázka je, či táto trieda má nejakú štruktúru, ktorá by charakterizovala triedu ako celok a ktorá by odzrkadlovala požiadavky matematickej štatistiky. Potreba takejto štruktúry, ktorá by nebola púhym reprodukovaním štruktúry teórie pravdepodobnosti, sa prejavila už v klasických prácach R. A. Fishera v 20. rokoch nášho storočia. Fisher zaviedol napríklad dva dôležité pojmy, ktoré nemajú analógiu v teórii pravdepodobnosti:

a) funkciu vierohodnosti

$$\theta \in \Theta \mapsto p(y \mid \theta) \in R^1$$

a z nej vyplývajúcu metódu odhadovania parametrov nazvanú metóda maxima vierohodnosti,

b) informačnú maticu, ktorá je definovaná ako kovariančná matica náhodného vektora

$$\left(\frac{\partial \ln p(y \mid \theta)}{\partial \theta_1}, \ldots, \frac{\partial \ln p(y \mid \theta)}{\partial \theta_m}\right).$$

Táto informačná matica je multivariantnou mierou množstva informácie o parametroch θ, získanej pozorovaním vektora y.

Ak sa z pohľadu súčasných teórií spätne dívame na tieto a podobné Fisherove koncepcie, môžeme povedať, že jeho nazeranie na matematickú štatistiku bolo do určitej miery diferenciálne-geometrické. Aj keď Fisherove výsledky boli vždy veľmi uznávané, tento geometrický aspekt nebol zdôrazňovaný (ani samotným Fisherom). Geometrická charakterizácia modelov matematickej štatistiky bývala aj pozdejšie ojedinelá a používala sa len v jednoduchých alebo špeciálnych prípadoch. Niektoré práce možno ovšem považovať v tomto smere za priekopnícke, pretože sa pokúšali o geometrický prístup ku všeobecným štatistickým problémom. Patria sem napr. práce C. R. Raa [1], N. N. Čencova [2] a B. Efrona [3, 4].

Rao upozornil na to, že Fisherovu informačnú maticu možno geometricky interpretovať ako tzv. metrický tenzor v parametrickom priestore Θ. N. N. Čencov dokázal, že takýto metrický tenzor je jediný, ak má byť ekvivariantný v istom štatistickom zmysle. Súčasne ukázal, že štatistické modely možno geometricky charakterizovať aj pomocou jednoparametrickej triedy tzv. afinných konekcií. Nakoniec Efron ukázal, že tzv. exponenciálne triedy rozdelení pravdepodobnosti majú jednoduchú geometrickú štruktúru, ktorá sa zvlášť prejavuje pri odhadovaní metódou maxima vierohodnosti. S malým časovým odstupom vzniklo niekoľko nezávislých stredísk výskumu geometrických prístupov ke matematickej štatistike. Možno tu uviesť práce Batesa, Wattsa a Hamiltona (napr. [5, 6]), Barndorffa-Nielsena (napr. [7]), Atkinsona a Mitchellovej [8],

Reida [9], Skovgaardovej [10] a hlavne Amariho ([11, 12]). Na Matematickom ústave SAV sa od r. 1981 využívajú geometrické metódy na vyšetrovanie neasymptotických vlastností nelineárnych štatistických modelov (napr. [13–15] a [21]).

Treba ešte poznamenať, že v prácach Amariho je priamy súvis s niektorými výsledkami teórie informácie (prezentovanými napr. v knihe I. Vajdu [16]).

V ďalších častiach článku chcem naznačiť, v čom spočíva geometrický prístup ku štatistike. Genézu tohto prístupu je vhodné sledovať počínajúc od najjednoduchšieho a súčasne najviac používaného štatistického modelu, od lineárneho regresného modelu [17].

2. Lineárny regresný model

Aby sme získali konkrétnejšiu predstavu takéhoto modelu, uvažujme ilustračný príklad: Na skúšobnej dráhe automobilky testujú nové automobily. V časovom rozpätí od 0 do 10 sekúnd sa automobil pohybuje zrýchlene (rozbieha sa). Pri konštantnej sile motora a so zanedbaním odporu vzduchu možno jeho polohu $s(t)$ v čase t vyjadriť vzťahom známym zo stredoškolskej fyziky

$$s(t) = vt + at^2/2$$

kde v je začiatočná rýchlosť (pre $t = 0$) a a je zrýchlenie. Parametre v a a sú neznáme. Poloha automobilu v časoch $t_1, \ldots, t_N$ sa meria prístrojmi a z nameraných výsledkov sa parametre v a a odhadujú. Merania sú sprevádzané náhodnými chybami, teda nameraná poloha y_i v čase t_i je

$$y_i = vt_i + at_i^2/2 + \varepsilon_i\,; \quad (i = 1, \ldots, N)\,, \tag{1}$$

kde $\varepsilon_1, \ldots, \varepsilon_N$ sú chyby merania, ktoré možno považovať za náhodné, nezávislé, s nulovými strednými hodnotami a s konštantnými disperziami.

Neznáme parametre v a a je vhodné odhadovať všeobecne používanou metódou najmenších štvorcov, t. j. riešiť úlohu

$$(\hat{v}, \hat{a}) = \arg\min_{v,a} \sum_{i=1}^{N} [y_i - (vt_i + at_i^2/2)]^2\,. \tag{2}$$

Úlohu (2) môžeme ľahko „geometrizovať": V N-rozmernom euklidovskom priestore E^N vytvoríme 2-rozmernú rovinu parametrizovanú parametrami v a a:

$$\mathscr{L} := \{z: z \in E^N,\ z_i = vt_i + at_i^2/2; (v, a) \in R^2,\ i = 1, \ldots, N\}\,.$$

Suma na pravej strane v (2) je druhá mocnina euklidovskej vzdialenosti vektora $y = (y_1, \ldots, y_N)^T$ od bodu roviny $\mathscr{L}$ určeného parametrami v a a. Pritom odhady $\hat{v}$ a $\hat{a}$ zodpovedajú bodu, ktorý je najbližšie ku y. Z elementárnej geometrie vieme, že takýto bod dostaneme ortogonálnou projekciou bodu y na $\mathscr{L}$. Ortogonálnu projekciu bodu y na $\mathscr{L}$ môžeme ovšem vyjadriť aj explicitne, v maticovom tvare. Za týmto účelom model (1) preformulujeme do všeobecnejšieho tvaru

$$y = X\theta + \varepsilon\,, \tag{3}$$

kde $y = (y_1, \ldots, y_N)^T$ je vektor nameraných údajov, $\theta = (\theta_1, \ldots, \theta_m)^T$ je vektor nezná-

mych parametrov, X je matica známych koeficientov. V tejto symbolike je

$$\mathscr{L} = \{X\theta : \theta \in R^m\} .$$

Keďže chyby merania ε_i majú nulové stredné hodnoty, rovina $\mathscr{L}$ je vlastne množina možných (ale neznámych) stredných hodnôt vektora y a geometricky reprezentuje našu neinformovanosť o meranom objekte. Ľahko sa možno presvedčiť, že ortogonálny projektor P na rovinu $\mathscr{L}$ je matica

$$P := X(X^TX)^{-1}X^T \tag{4}$$

(platí: $PP = P$, $P = P^T$, $Pz \in \mathscr{L}$ pre každé $z \in R^N$). Teda bod roviny $\mathscr{L}$, ktorý je najbližšie ku bodu y sa rovná

$$X\hat{\theta} = X(X^TX)^{-1}X^Ty .$$

To ale znamená, že hľadaný odhad má tvar

$$\hat{\theta} = (X^TX)^{-1}X^Ty , \tag{5}$$

čo sa zhoduje so vzorcom známym zo štatistiky ([18], kap. VII. 1).

Poznamenávame, že použité vzťahy sú správne, len ak matica X je plnej hodnosti. Uvedený geometrický prístup však nevyžaduje tento predpoklad, pretože je v podstate neparametrický (t.j. nezávisí od spôsobu parametrizovania roviny $\mathscr{L}$).

V modeli (3) sa výberový priestor (= množina možných hodnôt vektora y) zhoduje s R^N, presnejšie povedane zhoduje sa s euklidovským priestorom E^N, pretože výraz, ktorý minimalizujeme v (2), je štvorec euklidovskej normy vektora $y - X\theta$. Akú štatistickú interpretáciu má euklidovská štruktúra výberového priestoru? Formulujme otázku trochu ináč: Čo by sa stalo, keby sme pozmenili normu v R^N, t.j. keby sme namiesto rovnice (2) riešili rovnicu

$$\hat{\theta} := \arg\min_{\theta \in R^m} [y - X\theta]^T \Sigma^{-1} [y - X\theta] , \tag{6}$$

kde Σ je daná pozitívne definitná matica? Takáto modifikácia metódy najmenších štvorcov sa v praxi používa. V prípade, že matica Σ je diagonálna, nazýva sa váženou metódou najmenších štvorcov. Otázkou je, ktorá voľba matice Σ je najvhodnejšia, teda aká geometrická štruktúra výberového priestoru zodpovedá požiadavkám štatistiky? Odpoveď dáva klasická Gauss-Markovova veta, ktorá hovorí, že spomedzi všetkých odhadov typu (6) najpresnejšie odhady dostaneme, keď matica Σ je úmerná kovariančnej matici náhodného vektora y. Pokiaľ kovariančná matica vektora y nezávisí od θ (a to sa v lineárnych regresných modeloch predpokladá), definuje normu a tým aj geometriu vo výberovom priestore.

Ďalšou zaujímavou otázkou je, čo je geometrickým obrazom presnosti odhadu $\hat{\theta}$, resp. aká je geometrická štruktúra parametrického priestoru R^m z hľadiska tejto presnosti? Uvažujme dva body $\bar{\theta}$ a θ^* z parametrického priestoru. Ich štatistická rozlíšiteľnosť je daná štatistickou rozlíšiteľnosťou korešpondujúcich stredných hodnôt vektora y, t.j. rozlíšiteľnosťou bodov $X\bar{\theta}$ a $X\theta^*$. V zmysle vyššie povedaného táto rozlíšiteľnosť je určená normou

$$([X\bar{\theta} - X\theta^*]^T \Sigma^{-1} [X\bar{\theta} - X\theta^*])^{1/2}$$

kde Σ je kovariančná matica vektora y. Posledný výraz môžme ľahko napísať do tvaru

$$([\bar{\theta} - \theta^*]^T M[\bar{\theta} - \theta^*])^{1/2}, \tag{7}$$

kde matica M sa rovná

$$M = X^T \Sigma^{-1} X. \tag{8}$$

Dá sa ukázať, že v prípade, že chyby merania $\varepsilon_1, \ldots, \varepsilon_N$ v modeli (3) majú gaussovské rozdelenie pravdepodobnosti, matica M je proste Fisherova informačná matica definovaná v prvej časti článku. To ale znamená, že odpoveď na vyššie položenú otázku znie: parametrický priestor je lineárny normovaný priestor s normou (7), ktorá je určená informačnou maticou daného lineárneho regresného modelu. (Poznamenávame, že norma (7) sa v aplikáciách využíva napr. v diskriminačnej analýze.)

Naviac je známe, že kovariančná matica vektora $\hat{\theta}$ sa rovná matici M^{-1}, to znamená, že disperzie odhadov jednotlivých parametrov $\theta_1, \ldots, \theta_m$ sa rovnajú diagonálnym prvkom matice M^{-1}. (Pre $\Sigma = I$ vyplýva to priamo zo vzťahu (5).) V lineárnom modeli je teda evidentné, že matica M si zaslúži názov „informačná matica".

Zhrnieme úvahy tohto odseku: geometria lineárneho regresného modelu je geometria vhodne volených euklidovských priestorov, ktoré tento model reprezentujú (výberový priestor, parametrický priestor, priestor $\mathscr{L}$ stredných hodnôt). Dimenzie týchto priestorov a ich normy reprezentujú neurčitosť našich vedomostí o modelovanom objekte.

3. Gaussovský nelineárny regresný model

Kvôli konkrétnosti uvažujme opäť ilustračný príklad. Experimentátor meria výchylky tlmene kmitajúcej sústavy v časoch $t_1, \ldots, t_N$. Výchylka sústavy v čase t sa rovná

$$K \mathrm{e}^{-\beta t} \sin(\omega t + \varphi),$$

kde $K, \beta, \omega, \varphi$ sú fyzikálne významné parametre (amplitúda, tlmenie, frekvencia a fáza). Ich hodnoty sú neznáme a experimentátor ich má určiť na základe N meraní. Výsledky meraní sú

$$y_i = K \mathrm{e}^{-\beta t_i} \sin(\omega t_i + \varphi) + \varepsilon_i; \quad (i = 1, \ldots, N), \tag{9}$$

kde ε_i je opäť náhodná chyba i-tého merania. Množina hypoteticky možných stredných hodnôt vektora $y = (y_1, \ldots, y_N)^T$ sa rovná

$$\mathscr{E} := \{z: z \in R^N, z_i = K \mathrm{e}^{-\beta t_i} \sin(\omega t_i + \varphi), K > 0, \beta > 0, \omega > 0, \varphi \in (-\pi. \pi),$$
$$i = 1, \ldots, N\}.$$

Na rozdiel od podobnej množiny $\mathscr{L}$, ktorú sme uvažovali v lineárnom modeli, množina $\mathscr{E}$ vytvára zakrivenú plochu vnorenú do E^N.

Je vhodné opäť prejsť ku všeobecnej symbolike. Namiesto (9) píšeme

$$y = \eta(\theta) + \varepsilon, \tag{10}$$

kde η: $\Theta \mapsto R^N$ je zobrazenie definované na otvorenej množine $\Theta \subset R^m$, ktoré má

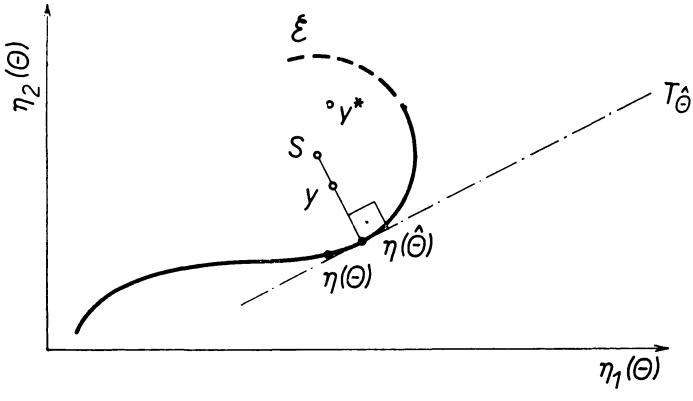
spojité druhé parciálne derivácie $\partial^2\eta(\theta)/\partial\theta_i\,\partial\theta_j$. Tu θ je vektor neznámych parametrov a ε je vektor chýb, o ktorom budeme predpokladať, že má gaussovské rozdelenie. Množina stredných hodnôt ε má v novej symbolike tvar

$$\mathscr{E} = \{\eta(\theta): \theta \in \Theta\}\,. \tag{11}$$

Podobne ako v lineárnom modeli, používame odhad metódou najmenších štvorcov, ktorý je definovaný rovnicou

$$\hat{\theta} = \arg\min_{\theta\in\Theta} [y - \eta(\theta)]^T [y - \eta(\theta)] \tag{12}$$

(pokiaľ existuje). Odhad $\hat{\theta}$ sa nedá tak jednoducho vypočítať ako v lineárnom modeli. Geometricky ide ovšem opäť o minimalizáciu vzdialenosti bodu y od plochy $\mathscr{E}$, a teda o ortogonálnu projekciu (obr. 1 pre prípad $m = 1$, $N = 2$).



Obr. 1.

Geometrické úvahy o nelineárnom regresnom modeli je vhodné začať dotykovou rovinou ku ploche $\mathscr{E}$ v nejakom bode θ^0. Táto dotyková rovina je množina

$$T_{\theta^0} := \left\{\eta(\theta^0) + \sum_{i=1}^{m} \frac{\partial\eta(\theta^0)}{\partial\theta_i}\,\tau_i\,;\ (\tau_1, \ldots, \tau_m) \in R^m\right\} \tag{13}$$

(pozri obr. 1 pre $\theta^0 = \hat{\theta}$).

Nahraďme plochu $\mathscr{E}$ jej dotykovou rovinou T_{θ^0}. V štatistickej interpretácii to znamená, že namiesto nelineárneho modelu (10) uvažujeme model

$$y - \eta(\theta^0) = \sum_{i=1}^{m} \frac{\partial\eta(\theta^0)}{\partial\theta_i}\,\tau_1 + \varepsilon_i\,;\quad (i = 1, \ldots, N)\,, \tag{14}$$

v ktorom neznámymi parametrami sú $\tau_1, \ldots, \tau_m$. Model (14) je lineárnou aproximáciou modelu (10); dotyková rovina teda definuje takúto aproximáciu. Podobné aproximácie sa v aplikáciách často používajú, ak disperzie chýb merania sú malé.

Postúpme kúsok ďalej v geometrických predstavách o dotykových rovinách ku ploche $\mathscr{E}$. Najprv vypočítajme odhad parametra $\hat{\tau}$ v linearizovanom modeli (14). Označme

$\theta^1 = \theta^0 + \hat{\tau}$. V bode θ^1 zostrojme novú dotykovú rovinu T_{θ^1} a v odpovedajúcom lineárnom aproximatívnom modeli vypočítajme odhad parametrov τ a z neho hodnotu θ^2, potom zostrojíme T_{θ^2} atď. Táto myšlienka je základom Gaussovej-Newtonovej iteračnej metódy výpočtu nelineárnych odhadov metódou najmenších štvorcov [19]. Hľadaný odhad $\hat{\theta}$ je totiž pevným bodom opísanej iteračnej procedúry; ak totiž v (14) položíme $\theta^0 = \hat{\theta}$, dostaneme $\hat{\tau} = 0$, teda dotyková rovina $T_{\hat{\theta}}$ je invariantná na procedúru.

Využitie dotykových rovín ku $\mathscr{E}$ je hlbšie v nasledujúcej úvahe. V lineárnom modeli (14) môžeme pomerne jednoducho zostrojiť tzv. oblasť spoľahlivosti pre parameter θ. Oblasť spoľahlivosti je taká podmnožina parametrického priestoru Θ, ktorá závisí od pozorovaného vektora y a ktorá pokrýva neznámu skutočnú hodnotu vektora parametrov θ s predpísanou pravdepodobnosťou (tzv. spoľahlivosťou). Označme $\mathscr{O}_{\theta^0}(y)$ oblasť spoľahlivosti v modeli (14). Model (14) je len aprocimáciou modelu (10), preto oblasť $\mathscr{O}_{\theta^0}(y)$ je len približná, jej spoľahlivosť nie je presne určená. Vytvorme preto množinu

$$V(y) := \{\theta^0 : \theta^0 \in \Theta, \theta^0 \in \mathscr{O}_{\theta^0}(y)\}.$$

Teda do $V(y)$ patria tie vektory θ^0, ktoré patria do „svojich" aproximatívnych oblastí $\mathscr{O}_{\theta^0}(y)$. Je pozoruhodné, že množina $V(y)$ je oblasť spoľahlivosti v pôvodnom nelineárnom modeli (10), a to oblasť presná, s presne určenou spoľahlivosťou. (Uvedená konštrukcia je geometrickou interpretáciou konštrukcie oblastí spoľahlivosti uvedenej v [20].) Čím je to, že pomocou aproximačných medzikonštrukcií v dotykových rovinách dostávame nakoniec presný výsledok? Je to zrejme preto, že uvažujeme množinu *všetkých* dotykových rovín $\{T_{\theta^0} : \theta^0 \in \Theta\}$, teda to, čo sa v diferenciálnej geometrii nazýva dotykový priestor plochy $\mathscr{E}$.

Uvažujme teraz na chvíľu namiesto lineárnej aproximácie (14) kvadratickú aproximáciu modelu (10). Ináč povedané, namiesto lineárnej časti Taylorovho rozvoja funkcie $\eta(\theta)$, ktorá vystupuje v (14), použime kvadratickú časť tohto rozvoja. Dostaneme takto množinu

$$K_{\theta^0} := \left\{ \eta(\theta^0) + \sum_{i=1}^{m} \frac{\partial \eta(\theta^0)}{\partial \theta_i} \tau_i + \frac{1}{2} \sum_{i,j=1}^{m} \frac{\partial^2 \eta(\theta^0)}{\partial \theta_i \, \partial \theta_j} \tau_i \tau_j : (\tau_1, \ldots, \tau_m) \in R^m \right\}.$$

Geometrická charakterizácia plochy $\mathscr{E}$ pomocou množiny K_{θ^0} nie je však veľmi vhodná, pretože na rozdiel od množiny T_{θ^0} množina K_{θ^0} závisí od spôsobu akým je parametrizovaná plocha $\mathscr{E}$. Ináč povedané, na rozdiel od dotyčnice T_{θ^0}, množinu K_{θ^0} nemôžeme nakresliť do obrázku 1, pokiaľ nemáme špecifikovanú parametrizáciu krivky $\mathscr{E}$. Geometrická charakterizácia kvadratických vlastností zakrivenej plochy je dosť obtiažna a diferenciálna geometria musela vyvinúť dômyselné pojmy, aby takúto charakterizáciu vykonala. Sú to v prvom rade rôzne definície krivostí. Aj tieto definície sa ukázali ako veľmi užitočné pre matematickú štatistiku. Pokúsime sa aspoň naznačiť toto využitie krivostí.

V prvom rade porovnanie polomerov krivosti plochy $\mathscr{E}$ s disperziami chýb merania udáva medze použiteľnosti nelineárneho regresného modelu (10). Odhadovanie parametrov v modeli (10) je totiž korektné, len ak disperzie chýb merania sú menšie ako najmenší polomer krivosti plochy $\mathscr{E}$. V opačnom prípade môže odhad (12) strácať zmysel a viesť ku scestným výsledkom. Situácia je naznačená na obr. 1, kde bod S je

stred krivosti krivky $\mathscr{E}$ v bode θ (θ = skutočná hodnota parametra). Ak sa pozorovaný vektor y^* dostane za stred krivosti, jeho projekcia padne do čiarkovanej časti krivky $\mathscr{E}$, ktorá je vzdialená od skutočnej strednej hodnoty vektora y.

Zložitejšie je nájdenie priamej súvislosti medzi krivosťami a tzv. asymptotickou efektívnosťou druhého rádu u odhadov. (Súvis je vysvetlený v článku [3] a v monografii [12]).

Autor tohto článku sa stretol s výhodným využitím polomerov krivosti pri hľadaní aproximatívneho vzťahu pre hustotu pravdepodobnosti odhadu (12) geometrickými metódami [14]. Rozdiel medzi aproximatívnou a presnou hustotou možno ohraničiť funkciou, ktorej premennou je najmenší polomer krivosti plochy $\mathscr{E}$.

Autori článku [5] zasa s výhodou používajú krivosti na to, aby charakterizovali medze použiteľnosti metód lineárnej štatistickej inferencie na nelineárny model (10).

Iné geometrické prostriedky boli využité v práci [13] pri vyšetrovaní jednoznačnosti odhadu (12). Šlo hlavne o využitie tzv. Sardovej vety z diferenciálnej geometrie a isté úvahy o dimenziách. Ide však o technické detaily, ktoré sa nehodia do nášho výkladu.

4. Zakrivené exponenciálne triedy rozdelení pravdepodobnosti

Každý, kto sa aspoň trochu zaoberal štatistikou, sa stretol s pojmom „histogram". Histogram je graf, ktorý rýchlo a prehľadne charakterizuje náhodnosť pozorovanej náhodnej veličiny. Histogram konštruujeme takto: vykonáme r nezávislých pozorovaní nejakej náhodnej veličiny ζ, ktorá môže nadobudnúť hodnoty z nejakého intervalu I. Interval I rozložíme na neprekrývajúce sa podintervaly $I_1, \ldots, I_{k+1}$. Označíme r_i počet, pozorovaní veličiny ζ, výsledky ktorých spadajú do I_i. Histogram je graf funkcie, ktorá na intervale I_i nadobúda konštantnú hodnotu r_i.

Získanie histogramu znamená často len prvé predspracovanie pozorovaných dát a histogram je východiskom pre ďalšiu štatistickú inferenciu. Veličiny $r_1, \ldots, r_{k+1}$ sú náhodné, a ako je dobre známe, kombinatorickými úvahami môžme odvodiť, že pravdepodobnosť hodnôt $r_1, \ldots, r_{k+1}$ sa rovná

$$P(r_1, \ldots, r_{k+1} \mid p_1, \ldots, p_{k+1}) = \frac{r!}{r_1! \ldots r_{k+1}!} p_1^{r_1} \ldots p_{k+1}^{r_{k+1}}, \tag{15}$$

kde p_i je pravdepodobnosť toho, že hodnota veličiny ζ je z intervalu I_i pri jednom pozorovaní. Ak nepoznáme rozdelenie veličiny ζ, vystupujú čísla p_i ako neznáme parametre, ktoré sú ovšem viazané podmienkou

$$\sum_{j=1}^{k+1} p_j = 1 .$$

Podobne platí:

$$\sum_{j=1}^{k+1} r_j = r .$$

Je preto rozumné reparametrizovať rozdelenie pravdepodobnosti (15). Nové parametre $\gamma_1, \ldots, \gamma_m$ môžu byť napríklad

$$\gamma_i = \ln p_i/p_{k+1}$$

a nové náhodné veličiny

$$y_i = r_i\,; \quad (i = 1, \ldots, k)\,.$$

Z (15) ľahko dostaneme, že

$$P(y \mid \gamma) = \frac{r!}{y_1! \ldots y_k! \,(r - \sum_j y_j)!} \exp\{\sum_j y_j \gamma_j\} \,(1 + \sum_j \mathrm{e}^{\gamma_j})^{-r}\,. \tag{16}$$

Zjednodušene môžeme (16) zapísať v tvare

$$P(y \mid \gamma(\theta)) = a(y)\,\mathrm{e}^{y^T\gamma - \varkappa(\gamma)}\,; \quad (\gamma \in \Gamma)\,, \tag{17}$$

kde $a(y)$, $\varkappa(\gamma)$ sú vhodne zvolené funkcie a kde množina Γ možných hodnôt vektora parametrov γ (= parametrický priestor) je otvorená podmnožina R^k. Triedy rozdelení pravdepodobností ktoré možno zapísať v tvare (17) sa nazývajú (lineárne) exponenciálne triedy. To sú veľmi dôležité triedy rozdelení pravdepodobnosti; patria sem mnohé štatistické modely a uvedený príklad analýzy histogramu je len jeden z mnohých prípadov použitia takýchto tried. História ich vzniku siaha ešte do raných období štatistickej fyziky, kde sa rozdeleniami pravdepodobnosti typu (17) opisovali stavy plynov; parametre $\gamma_1, \ldots, \gamma_m$ pritom zodpovedali termodynamickým veličinám, ako sú teplota, tlak a podobne.

V matematickej štatistike nazývame parametre $\gamma_1, \ldots, \gamma_m$ kanonickými parametrami triedy (17).

V niektorých prípadoch nie sú parametre $\gamma_1, \ldots, \gamma_k$ celkom neznáme, ale sú samy funkciami menšieho počtu iných neznámych parametrov, ktoré označíme $\theta_1, \ldots, \theta_m$. To znamená, že namiesto triedy (17) máme triedu

$$P(y \mid \gamma(\theta))\,; \quad (\theta \in \Theta)\,, \tag{18}$$

ktorá je podtriedou triedy (17). V prípade, že množina Θ je otvorená v R^m a že funkcie $\gamma_1(\theta), \ldots, \gamma_k(\theta)$ majú spojité druhé parciálne derivácie, nazývame triedu (18) zakrivenou exponenciálnou triedou.

Vo vyššie uvedenom príklade histogramu, dostaneme zakrivenú exponenciálnu triedu, ak je rozdelenie pravdepodobnosti pozorovanej veličiny ζ čiastočne známe už pred pozorovaním.

Iným príkladom zakrivenej exponenciálnej triedy je nelineárny regresný model (10). Príkladom (lineárnej) exponenciálnej triedy je lineárna regresia uvedená v odseku 2.

Zakrivené exponenciálne triedy majú zaujímave geometrické vlastnosti. Súvisia s metódou odhadovania parametrov, ktorá sa v takýchto triedach najčastejšie používa. Je to odhad maxima vierohodnosti definovaný vzťahom

$$\hat{\theta} := \arg\max_{\theta \in \Theta} \ln P(y \mid \gamma(\theta))\,. \tag{19}$$

Intuitívne zdôvodnenie odhadu (19) je zrejmé. Za „najvierohodnejšiu" považujeme tú hodnotu parametra θ, pre ktorú je najpravdepodobnejší práve obdržaný výsledok y.

Možno takto definovaný odhad dať do súvisu s nejakou geometriou výberového priestoru, tak ako sa to podarilo v odsekoch 2 a 3? Zodpovedá odhad (19) minimalizácii

nejakej vzdialenosti bodu y napr. od plochy $\{\gamma(\theta): \theta \in \Theta\}$? Súvisí odhad (19) s ortogonálnou projekciou na túto plochu?

Niekoľko technických detailov a predpokladov uľahčí zodpovedanie týchto otázok. Predpoklady sa týkajú „nezakrivenej" triedy (17), do ktorej je vnorená zakrivená trieda (18).

a) Predpokladáme, že kanonický parametrický priestor Γ je otvorená podmnožina R^k a že funkcia $\varkappa(\gamma)$ je dva razy spojite diferencovateľná.

b) Derivovaním evidentnej rovnosti

$$\sum_y P(y \mid \gamma) = 1 ,$$

kde $P(y \mid \gamma)$ je určené vzťahom (17), sa môžeme presvedčiť, že stredná hodnota vektora y je

$$E_\gamma(y) = \left(\frac{\partial \varkappa(\gamma)}{\partial \gamma_1}, \ldots, \frac{\partial \varkappa(\gamma)}{\partial \gamma_k}\right)^T . \tag{20}$$

Budeme predpokladať, že zobrazenie

$$\gamma \mapsto E_\gamma(y)$$

definované vzťahom (20) je injekcia do R^k a že množina

$$H := \{E_\gamma(y): \gamma \in \Gamma\} \tag{21}$$

pokrýva celý výberový priestor vektora y.

V uvedenom príklade vyšetrovania histogramu sú uvedené predpoklady splnené. Podobne všetky predpoklady sú splnené aj v nelineárnom regresnom modeli (10).

c) Nakoniec jedno označenie. V triede (18) budeme písať

$$\eta(\theta) := E_{\gamma(\theta)}(y) ,$$

teda strednú hodnotu vektora y budeme označovať rovnako ako v predchádzajúcich odsekoch článku.

Začnime rozborom vhodnej „vzdialenosti" vo výberovom priestore. Teória informácie sa už dávnejšie dôkladne zaoberá hľadaním štatisticky interpretovateľných „vzdialeností" ľubovoľných dvoch rozdelení pravdepodobnosti (pozri [16]). Vo všeobecnosti sa ukázalo, že takéto „vzdialenosti" sa nedajú definovať ako metriky; zaužíval sa pre ne názov divergencie (my však budeme i naďalej hovoriť aj o „vzdialenostiach"). Divergencie sú síce nezáporné a sú nulové práve vtedy, keď rozdelenia pravdepodobnosti, ktorých „vzdialenosť" určujeme, sú totožné. Avšak divergencie nemusia byť ani symetrické. To znamená, že „vzdialenosť" nejakého rozdelenia P od iného rozdelenia $\bar{P}$ môže byť iná ako „vzdialenosť" $\bar{P}$ od P.

Známou divergenciou je tzv. I-divergencia. Jej pôvod je v známej Shannonovej miere informácie a má svoj pôvod ešte v Bolzmanovej štatistickej fyzike z minulého storočia.

I-divergencia je definovaná vzťahom

$$I(\gamma, \gamma^*) = \sum_\gamma \left[\ln \frac{P(y \mid \gamma)}{P(y \mid \gamma^*)}\right] P(y \mid \gamma) . \tag{22}$$

Vo všeobecnosti je výpočet I-divergencie zložitý. Ale v exponenciálnej triede (17) platí, ako sa možno ľahko presvedčiť dosadením (17) do (22), jednoduchý vzťah

$$I(\gamma, \gamma^*) = (\gamma - \gamma^*)^T \eta - \varkappa(\gamma) + \varkappa(\gamma^*), \tag{23}$$

kde sme označili $\eta := E_\gamma(y)$. V tomto výraze pre I-divergenciu svorne vystupujú oba možné spôsoby parametrizácie triedy (17): pomocou kanonických parametrov γ a pomocou strednej hodnoty η.

Dvojica parametrizácií γ a η vystupuje „duálne" v celej geometrickej teórii exponenciálnych tried. Sú to síce dve rôzne parametrizácie a každá z nich stačí na označenie prvkov triedy (17), avšak v mnohých výrazoch je výhodné používať obe parametrizácie súčasne, spoločne, tak ako v (23). Je pozoruhodné, že v lineárnom alebo v nelineárnom regresnom modeli (10) sa obe parametrizácie stotožnia a pre I-divergenciu platí

$$I(\gamma, \gamma^*) = \tfrac{1}{2}\|\gamma - \gamma^*\|^2 .$$

Z uvedeného vyplýva, že v triede (17) môžeme definovať funkciu $J(\eta, \eta^*)$ vzťahom

$$J(\eta, \eta^*) = I(\gamma, \gamma^*),$$

t.j. vlastne výrazom (23). Funkcia $J(\eta, \eta^*)$ definuje „vzdialenosť" v priestore H, a teda podľa vyššie uvedeného predpokladu b definuje aj „vzdialenosť" vo výberovom priestore. Teda „vzdialenosť" bodu y od plochy $\mathscr{E} := \{\eta(\theta): \theta \in \Theta\}$ sa rovná

$$J(y, \eta(\theta)) = [\gamma_y - \gamma(\theta)]^T y - \varkappa[\gamma_y] + \varkappa[\gamma(\theta)]. \tag{24}$$

Pritom vektor γ_y je riešením rovnice (20), kde za $E_\gamma(y)$ dosadíme priamo vektor y. To však poznamenávame len kvôli úplnosti. Pre nás je dôležité, že pomocou vzťahov (17) a (23) sa môžeme ľahko presvedčiť, že súčet

$$\ln P(y \mid \gamma(\theta)) + J(y, \eta(\theta))$$

nie je závislý od vektora parametrov θ. Odtiaľ vyplýva, že odhad metódou maxima vierohodnosti môžeme ekvivalentne získať aj z rovnice

$$\hat{\theta} = \arg\min_{\theta \in \Theta} J(y, \eta(\theta)),$$

teda pomocou minimalizácie I-divergencie. I-divergencia je teda hľadaná „vzdialenosť", ktorá korešponduje s metódou maxima vierohodnosti a I-divergencie charakterizuje geometriu výberového priestoru.

Ako je to s hľadanou ortogonalitou vektora $y - \eta(\theta)$ na plochu $\mathscr{E}$? Keďže vo výberovom priestore nemáme metriku, ale iba I-divergenciu, ťažko hovoriť o ortogonalite dvoch vektorov v obvyklom zmysle. Určité vodítko ovšem poskytuje nasledujúca veľmi zaujímavá analógia Pythagorovej vety, ktorú môžeme jednoducho dokázať. Pre tri body γ, $\bar{\gamma}$, γ^* z množiny Γ totiž platí rovnosť

$$I(\gamma, \bar{\gamma}) + I(\bar{\gamma}, \gamma^*) = I(\gamma, \gamma^*),$$

akonáhle platí

$$(\gamma - \bar{\gamma})^T (\bar{\eta} - \eta^*) = 0 .$$

Túto rovnosť preveríme púhym dosadením výrazu (23). Ortogonalita je teda „spoločná" pre obe parametrizácie γ a η. Možno hovoriť napr. o priamkach v priestore Γ, ktoré sú kolmé na priamky v priestore H, ale nie o kolmých priamkach ležiacich v tom istom priestore. Preto sa dá očakávať, že v metóde maxima vierohodnosti nebude vektor $y - \eta(\theta)$ kolmý na plochu $\{\eta(\theta): \theta \in \Theta\}$, ale skôr na plochu $\{\gamma(\theta): \theta \in \Theta\}$. O tom sa ľahko presvedčíme úpravou rovníc

$$\frac{\partial}{\partial \theta_i} \ln P(y \mid \gamma(\theta))|_{\theta=\hat{\theta}} = 0\,; \quad (i = 1, \ldots, m)\,,$$

ktoré priamo vyplývajú z (19). Po dosadení podľa vzťahu (18) a po využití rovnosti (20) dostaneme nakoniec

$$[y - \eta(\hat{\theta})]^T \frac{\partial \gamma(\hat{\theta})}{\partial \theta_i} = 0\,; \quad (i = 1, \ldots, m)\,,$$

kde $\partial\gamma(\hat{\theta})/\partial\theta_1, \ldots, \partial\gamma(\hat{\theta})/\partial\theta_m$ sú zrejme dotykové vektory plochy $\{\gamma(\theta): \theta \in \Theta\}$.

Otázkou je, či možno charakterizovať geometriu výberového priestoru ešte ináč než pomocou I-divergencie, resp. vyššie definovanej ortogonality. Tento priestor nemá metriku. Diferenciálna geometria ovšem pozná spôsob popisu nemetrických priestorov, a to pomocou vnútorných (tzv. kovariantných) derivácií, alebo čo je ekvivalentné, pomocou tzv. afinných konekcií. My sa uspokojíme s konštatovaním, že takáto štruktúra je pre exponenciálne triedy vybudovaná v monografii [12].

Literatúra

[1] Rao, C. R.: *On the distance between two populations.* Sankhya *9* (1949), 246—248.

[2] Čencov, N. N.: *Statističeskije rešajuščije pravila i optimaľnyje vyvody.* Moskva, Nauka 1972.

[3] Efron, B.: *Defining the curvature of a statistical problem (with application to second order efficiency).* Ann. Statist. *3*, (1975), 1189—1242.

[4] Efron, B.: *The geometry of exponential families.* Ann. Statist. *6* (1978) 362—376.

[5] Bates, D. M., Watts, D. G.: *Relative curvature measures of non-linearity.* J. Roy. Statist. Soc. *B 40* (1980) 1—25.

[6] Bates, D. M., Hamilton, D. C., Watts, D. G.: *Calculation of intrinsic and parameter effects curvatures for nonlinear regression models.* Communications in Statistics *12* (1983) 469—477.

[7] Barndorff-Nielsen, O. E.: *Differential and integral geometry in statistical inference.* Research Report 106 (1984) Dept. Theor. Statist., Aarhus University.

[8] Atkinson, C., Mitchell, A. F. S.: *Rao's distance measure.* Sankhya *43* (1981) 345—365.

[9] Reid, N.: *Curvature and linear rank statistics.* Tech. Rep. N^0 83—13 (1983) Inst. App. Math. Statist., Univ. British Columbia.

[10] Skovgaard, L. T.: *A Riemannian geometry of the multivariate normal model.* Scand. J. Statist. *11* (1984), 211—222.

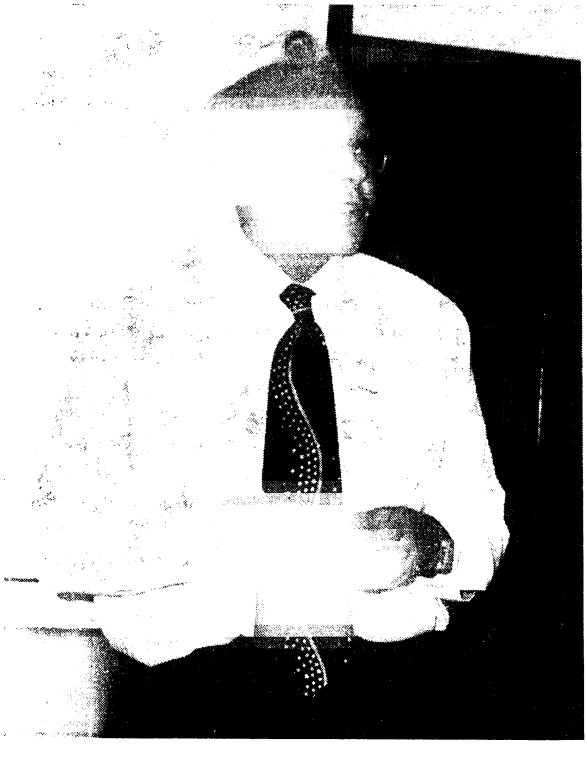
[11] Amari, S. I.: *Differential geometry of curved exponential families-curvatures and information loss.* Ann. Statist. *10* (1982) 357—385.

[12] Amari, S. I.: *Differential-Geomotrical Methods in Statistics.* Lecture Notes in Statistics N° 28, Springer: W. Berlin—Heidelberg, 1985.

[13] Pázman, A.: *Nonlinear least squares — uniqueness versus ambiguity.* Math. Operations. Statist., Ser. Statist. *15*, (1984) 323—336.

[14] PÁZMAN, A.: *On formulas for the distribution of nonlinear L. S. estimates.* Statistics *18* (1987) 3—15.
[15] PÁZMAN, A.: *On the uniqueness of the M. L. estimates in curved exponential families.* Kybernetika *22* (1986) 124—132.
[16] VAJDA, I.: *Teória informácie a štatistického rozhodovania.* ALFA: Bratislava 1982.
[17] KUBÁČEK, L.: *Základy teórie odhadu.* Veda: Bratislava 1983.
[18] ANDĚL, J.: *Matematická statistika.* SNTL/ALFA: Praha 1978.
[19] HARTLEY, H. O.: *The modified Gauss-Newton method for the fitting of non-linear regression functions by least squares.* Technometric *3* (1961) 269—280.
[20] SUNDARARAJ, N.: *A method for confidence regions for nonlinear models.* Austral. J. Statist. *20* (1978) 270—274.
[21] KOUTKOVÁ H.: *Odhady v modelu singulární nelineární regrese.* Kandidátska disertačná práca. Bratislava 1988.

Martin Černohorský očima svých žáků



Napsat cokoliv uceleného o člověku aktivit doc. Martina Černohorského je téměř vyloučeno. Proto se autoři medailónů většinou zaměřují jen na jednotlivé stránky jeho rozmanité činnosti, jako např. na semináře, které M. Černohorský s úspěchem organizuje.

Role opravdového pedagoga je však u něho do té míry dominantní, že se několik jeho žáků různého věku (léta 1952 až 1988) rozhodlo spojit své vzpomínky v mozaiku, kterou u příležitosti jeho pětašedesátých narozenin čtenářům předkládáme.

Redakce

Foto M. Lýčka

S Martinem Černohorským jsem přišel poprvé do styku jako student přírodovědecké fakulty UJEP v Brně při praktiku v roce 1952. Tehdy byl mladým asistentem. A již v té