

NAVRHOVANIE OPTIMÁLNYCH NELINEÁRNYCH REGRESNÝCH EXPERIMENTOV

Andrej PÁZMAN
MFF UKo, KPMŠt

Abstract: We present some features of experimental design in nonlinear regression models. In Section 2 is an outline of design theory in linear models. In the other sections we present successively the approach based on the Fisher information matrix, the sequential design, we explain briefly how to build optimality criteria from second order approximations of moments or entropy, or how to express optimality criteria based on the probability density of the estimator.

Резюме: Предложены некоторые свойства планирования экспериментов в нелинейной регрессионной модели. В части 2 в сжатом виде излагается теория планирования линейных экспериментов. В остальных частях постепенно обсуждаются подход использующий информационную матрицу Фишера, последовательное планирование эксперимента, дальше обсуждаются критерии оптимальности использующие аппроксимации второго порядка для моментов и энтропии, или использующие аппроксимацию плотности вероятности оценок параметров.

1 Úvod

Začnime motivačnými príkladmi

Príklad 1. Istý fyzik prišiel nedávno s takýmto problémom. Na jednom konci kovovej trubice sa zapáli plameň, ktorý sa postupne šíri do celej trubice. Do trubice je potrebné prevŕtať niekoľko (málo) otvorov tak, aby bolo možné sledovať, rýchlosť šírenia plameňa. Treba určiť polohu otvorov tak, aby závislosť $t(x)$ času objavenia sa plameňa od polohy x v trubici bola čo najlepšie aproximovaná polynómom 3. stupňa. Ide teda o navrhovanie experimentu pre optimálnu interpoláciu polynómu 3. stupňa.

Príklad 2. (Dĺžka Prístavného mostu v Bratislave.) Pri stavbe mostu cez Dunaj bolo potrebné presne zmerať vzdialenosť medzi dvomi bodmi ležiacimi na opačných brehoch rieky (tzv. vytýčenie osi mostu). Priame meranie tejto dĺžky nebolo možné, pretože požadovaná presnosť vyžadovala použitie optického (laserového) prístroja, ktorý neznášal reflexie od vodnej hladiny. Preto na každom brehu rieky boli vytýčené ďalšie dva body. V takto vytvorenej geodetickej sieti skladajúcej sa zo 6 bodoch bolo možné zmerať ľubovoľný uhol a ďalej každú vzdialenosť medzi bodmi ležiacimi na tej istej

strane rieky. Otázka znela: ktoré dĺžky a uhly zmerať, resp. ktoré merania opakovať a koľkokrát, aby pri danom celkovom počte meraní, hľadaná vzdialenosť bola určená čo najpresnejšie (nepriamo, pomocou známych vzťahov medzi dĺžkami strán a veľkosťami uhlov v trojuholníku).

Príklad 3. Kinetické reakcie v chemii. Kinetické reakcie bývajú často opisované nelineárnymi regresnými modelmi. Návrh experimentu tu spočíva vo výbere množstiev reagujúcich látok. Typickým je klasický príklad modelu Michaelisa a Mentena, kde množstvo enzymu x spôsobuje rýchlosť reakcie danú vzťahom

$$\frac{\theta_1 x}{\theta_2 + x}$$

, kde θ_1, θ_2 sú parametre, ktorých hodnoty treba odhadnúť z experimentu. V pokusoch treba optimálne voliť množstva enzymov x tak, aby presnosť odhadov parametrov bola čo najväčšia.

V súlade s uvedenými príkladmi je nasledujúca všeobecná formulácia úlohy navrhovania experimentov (pozri [11]).

Množina \mathcal{X} je množina možných pokusov. V každom pokuse $x \in \mathcal{X}$ možno pozorovať hodnotu reálnej náhodnej veličiny $y(x)$ a to tak, že pozorovania v rôznych pokusoch, alebo v opakovaných pokusoch, sú nezávislé.

Úlohou navrhovania experimentu je zvoliť N -ticu bodov

$$x_1, \dots, x_N$$

patriacich do množiny \mathcal{X} , tak, aby vektor nameraných údajov

$$y = (y(x_1), \dots, y(x_N))^T$$

obsahoval maximum požadovanej informácie. Uvedená N -tica sa nazýva **návrhom experimentu**. Číslo N je daný celkový počet pokusoch povolených v experimente. Body x_1, \dots, x_N nemusia byť rôzne, t.j. opakovania pokusov sú povolené.

Toto je klasická schéma poňatia experimentu a jeho návrhu. Modifikácie sú však možné. Napr.

- Namiesto počtu pokusov sa predpisujú celkové náklady na experiment. Podotýkame, že v lineárnom modeli to vedie ku nepodstatnej modifikácii teórie.

- V každom pokuse pozorujeme skupinu náhodných veličín, poprípade celý náhodný proces. Modifikácie teórie, ku ktorým takáto zmena vedie, sú zvládnuteľné podobnými prostriedkami ako v klasickom prípade (pozri [4]).

- Pozorovania vykonávané v rôznych pokusoch sú závislé, alebo opakovania pokusov nie sú dovolené. Takéto modifikácie vyžadujú úplne odlišné prístupy ku navrhovaniu experimentov i v lineárnych modeloch.

Vráťme sa k našej základnej schéme experimentu a jeho návrhu. Objavujú sa tu 3 základne otázky, ktoré zvlášť vyniknú, ak model je nelineárny.

- A) Ako definovať množstvo informácie získanej z experimentu.
- B) Ako predikovať túto informáciu ešte pred vykonaním experimentu.
- C) Ako vypočítať požadovaný návrh.

Ak chceme budovať teóriu navrhovania experimentov, treba najprv vedieť, aký model experimentu a aké metódy spracovania údajov budeme používať. Preto takúto teóriu možno budovať len pre dobré preskúmané modely a metódy inferencie. Prim tu hrá klasický lineárny regresný model spolu s metódou najmenších štvorcov. Pre tento prípad je teória tak dobre a elegantne vybudovaná, že by mala byť súčasťou základných kurzov.

2 O elegancii teórie navrhovania lineárnych experimentov

Model je definovaný vzťahmi

$$E_{\theta}[y(x)] = f^T(x)\theta, \quad \text{Var}[y(x)] = \sigma^2$$

platnými pre každé $x \in \mathcal{X}$ a pre každé $\theta \in R^m$. Vektor $f(x)$ je vektor známych koeficientov, vektor θ je vektor neznámych parametrov. Kvôli názornosti uvádzame, že v príklade 1 je $f^T(x) = (1, x, x^2, x^3)$ a neznámymi parametrami sú koeficienty polynomu.

Podľa klasickej Gaussovej-Markovovej vety "najlepšimi" odhadmi parametrov θ (t.j. nevychýlenými, a s minimálnou disperziou) sú odhady $\hat{\theta}$ získané metódou najmenších štvorcov (MNŠ). Navyac platí:

a) ľubovoľná lineárna funkcia parametrov tvaru $h^T\theta$ je odhadnuteľná práve vtedy, keď $h \in \mathcal{M}[M(x_1, \dots, x_N)]$. Tu

$$M(x_1, \dots, x_N) = \sum_{i=1}^N f(x_i)f^T(x_i)$$

a symbol \mathcal{M} označuje stĺpcový priestor príslušnej matice.

b) Pre disperziu odhadu tejto funkcie platí:

$$\text{Var}[h^T\hat{\theta}] = \sigma^2 h^T M^{-1}(x_1, \dots, x_N) h,$$

kde M^{-1} označuje g-inverziu matice M . Teda mieru množstva informácie získanej z experimentu určuje informačná matica $\sigma^{-2}M(x_1, \dots, x_N)$. Centrálnu úlohu tejto matice v lineárnom modeli potvrdzuje aj známe Rao-Cramerove dolné ohraničenie disperzie odhadov. Navyac, mnohé alternatívne odhady parametrov v modeli odvodzujú svoje asymptotické disperzie taktiež od informačnej matice. Tým je v podstate zodpovedaná otázka A) z úvodu. Ale ani otázka B) nespôsobuje ťažkosti v lineárnom modeli, pretože informačná

matica nezávisí od skutočnej hodnoty parametra θ a teda možno ju určiť ešte pred vykonaním experimentu.

V oblasti výpočtov optimálnych návrhov v lineárnych modeloch vznikol prielom, keď J.Kiefer ([8]) zaviedol tzv. návrhovú mieru. Ide o jednoduchú modifikáciu návrhu experimentu, spojenú ovšem s malým, ale veľmi účinným trikom. Prepíšeme informačnú maticu do tvaru

$$M(x_1, \dots, x_N) = NM(\xi)$$

kde $\xi(x)$ je definované ako relatívny počet opakovaní pokusu x (v danom návrhu experimentu) a kde

$$M(\xi) = \sum_{x \in \mathcal{X}} f(x)f^T(x)\xi(x)$$

Teda špeciálne platí, že $\xi(x) = 0$ ak sa pokus x nevyskytuje v návrhu. Je zrejmé, že ξ je pravdepodobnostná miera na \mathcal{X} . Nazýva sa návrhová miera prisluchajúca ku danému návrhu. Maticu $M(\xi)$ nazývame informačnou maticou návrhovej miery ξ . Pretože číslo N je pevné, o množstve informácie rozhoduje práve matica $M(\xi)$.

”Trik” je v tom, že **každú** pravdepodobnostnú mieru na množine \mathcal{X} , koncentrovanú do konečného počtu bodov, budeme považovať za návrhovú mieru. (Dobrá aproximácia každej takejto návrhovej miery nejakým návrhom je možná, ak počet meraní v experimente je veľmi veľký.) Zisk z takéhoto rozšírenia poňatia návrhu experimentu je výrazný, pretože rozšírenie umožňuje používať konvexné metódy (ich najvýraznejšie využitie možno nájsť v [16]). Množina všetkých možných návrhových mier (ale aj všetkých možných informačných matic) sa stáva konvexnou. Navyac, tzv. kritéria optimality návrhu experimentu, čo sú vhodne volené (t.j. štatisticky dobre interpretovateľné) funkcie informačnej matice, sú zväčša konvexnými funkciami. Takými sú napr. dobre známe kritérium D-optimality, $\Phi(M) = -\ln \det(M)$, kritérium A-optimality, $\Phi(M) = \text{tr}(M^{-1})$, kritérium G-optimality, $\Phi(M) = \max_{x \in \mathcal{X}} f^T(x)M^{-1}f(x)$, atd. (pozri [11], kap. IV.)

Uvedieme niekoľko elegantných a pritom závažných dôsledkov tejto konvexity. (Dôkazy možno nájsť napr. v [11], v tvrdeniach III.10, III.6, IV.27.)

Dôsledok 1. Každú návrhovú mieru možno nahradiť ekvivalentnou návrhovou mierou sústredenou najviac do $m(m+1)/2 + 1$ bodov množiny \mathcal{X} . (Ekvivalenciou sa rozumie zhoda informačných matic.)

Dôsledok 2. Ak $x \in \mathcal{X}$, avšak vektor $f(x)$ nie je na hranici množiny

$$\text{co}\{f(x) : x \in \mathcal{X}\} \cup \{-f(x) : x \in \mathcal{X}\},$$

tak pokus x nie je informatívny a treba ho vyradiť z množiny \mathcal{X} . Teda zaradením tohto pokusu do návrhu experimentu nemožno zlepšiť hodnotu žiadneho z hore uvedených kritérií optimality.

Poznamenávame, že v horeuvedenej množine symbol μ označuje konvexný obal množiny.

Dôsledok 3. Nech μ je návrhová miera a nech existuje gradient funkcie $\Phi(M)$ v bode $M = M(\mu)$. (Gradient tejto funkcie, označme ho $\nabla\Phi(M)$, je definovaný ako matica zložená z prvých derivácií funkcie Φ podľa prvkov matice M). Potom platí: návrhová miera μ je optimálna podľa kritéria Φ práve vtedy keď

$$\min_x f^T(x) \nabla\Phi[M(\mu)] f(x) = \text{tr} M(\mu) \nabla\Phi[M(\mu)]$$

Podotýkame, že poznáme jednoduché explicitné vzťahy pre vyjadrenie gradientu kritéria D-optimality, A-optimality, a pre mnohé ďalšie kritéria. Teda dôsledok 3 je dobre použiteľný pre praktické preverovanie rôznych návrhov experimentu, ktoré sa zdajú experimentátorovi výhodne z tohto ktorého dôvodu. Môžeme totiž rýchlo zistiť, či tieto návrhy sú blízke optimálnym návrhom (podľa rôznych kritérií optimality), alebo nie.

Všetky doterajšie úvahy sa týkali optimalizácie experimentu pre odhadovanie parametrov θ . Je však známe, že úloha odhadovania parametrov a úloha testovania samotného modelu sú úlohy komplementárne, teda návrh, ktorý je výborný na odhadovanie parametrov, môže byť veľmi zlý pre testovanie modelu. Tu však opäť, dôsledok 3 umožňuje zvoliť ad hoc návrh, ktorý je dosť dobrý pre testovanie, a posúdiť jeho efektívnosť vzhľadom na odhadovanie parametrov. Inou možnosťou je formulovať rôzne kompromisné kritéria optimality, umožňujúce ako testovanie tak ak efektívny odhad parametrov (pozri [9]).

Treba tiež zdôrazniť, že konvexnosť poskytuje dobre a elegantne prezentovateľné algoritmy pre výpočet optimálnych návrhových mier. Podrobnosti možno nájsť v [11], kap. V., prípadne najnovšie v [6].

3 Kritéria optimality a navrhovanie nelineárneho regresného experimentu

V nelineárnom modeli v každom pokuse $x \in \mathcal{X}$ pozorujeme náhodnú veličinu $y(x)$ so strednou hodnotou a disperziou

$$E_\theta[y(x)] = \eta(x, \theta), \quad \text{Var}[y(x)] = \sigma^2$$

kde vektor parametrov θ je opäť neznámy (a zvyčajne lokalizovaný v nejakej vopred danej množine Θ), a kde $\eta(\cdot, \cdot)$ je daná funkcia 2 krát spojitě diferencovateľná v oboch argumentoch. Pokiaľ neuvažujeme veľmi rozsiahle návrhy experimentov, je potrebné taktiež predpokladať, že pozorovaná veličina $y(x)$ je normálne rozdelená.

Zdalo by sa, že v teoretickej rovine prechod od lineárneho k nelineárnemu modelu je nepodstatný. Nie je to však tak, v nelineárnom modeli je obtiažne odpovedať už na otázky A) a B) kladené v úvode. Budeme sa zaoberať hlavne odpoveďami na tieto otázky, a odpoveď na otázku C) necháme zväčša otvorenú. Oprávňuje nás k tomu rýchly rozvoj výpočtových prostriedkov na počítačoch. To nám umožňuje uvažovať v ďalšom fixovaný návrh experimentu x_1, \dots, x_N , ktorý definujeme podobne ako v lineárnom modeli. Pokiaľ to nebude nevyhnutné,

NEBUDEME VYZNAČOVAŤ ZÁVISLOSŤ UVAŽOVANÝCH SYMBOLOV NA NÁVRHU EXPERIMENTU.

Budeme napr. používať symboly:

$y = (y(x_1), \dots, y(x_N))^T =$ vektor údajov

$\eta(\theta) = (\eta(x_1, \theta), \dots, \eta(x_N, \theta))^T =$ vektor stredných hodnôt

symbolom $f(y|\theta)$ označíme (normálnu) hustotu pravdepodobnosti náhodného vektora y . Informačná matica ma v nelineárnom modeli tvar

$$E_{\theta} \left\{ -\frac{\partial^2 \ln f(y|\theta)}{\partial \theta_k \partial \theta_l} \right\} = \sigma^{-2} M_{kl}(x_1, \dots, x_N; \theta)$$

kde maticu $M(x_1, \dots, x_N; \theta)$ dostaneme z matice $M(x_1, \dots, x_N)$ uvedenej v časti 2, ak vektor $f^T(x_i)$ nahradíme vektorom

$$\left(\frac{\partial \eta(x_i, \theta)}{\partial \theta_1}, \dots, \frac{\partial \eta(x_i, \theta)}{\partial \theta_N} \right)$$

Ako je známe, v nelineárnom experimente informačná matica má význam len v asymptotickom prípade. (Jej inverzia sa rovná asymptotickej variančnej matici MNS odhadu pre θ .) Navyiac, na rozdiel od lineárneho modelu, táto matica závisí od skutočnej hodnoty θ . Inak sa ale úplne zhoduje s informačnou maticou v lineárnom modeli. Platí teda tvrdenie:

Pokiaľ navrhujeme rozsiahle experimenty a tušíme, čomu by sa mala rovnať skutočná hodnota θ , môžeme v plnom rozsahu použiť výsledky známe z navrhovania lineárnych experimentov, pričom vektor $f^T(x)$ nahradíme vektorom parciálnych derivácií funkcie $\eta(x, \cdot)$ v bode θ .

Objavujú sa ovšem dve otázky:

a) čo robiť, keď dovolený rozsah návrhu nie je príliš veľký ,

b) ako obísť neznalosť skutočnej hodnoty θ . (Koniec koncom, cieľom navrhovaného experimentu je práve odhad vektora θ).

Otázkou a) sa budeme podrobne zaoberať v paragrafoch 3.3 a 3.4. Na otázku b) možno, zdá sa, odpovedať štyrmi spôsobmi. 1. Experiment navrhujeme sekvenčne (a to je zrejme najserioznejší spôsob, pokiaľ si ho môžeme dovoliť). 2. Namiesto odhadovania parametrov použijeme bayesovské metódy založené ovšem na znalosti apriorného rozdelenia a užitkovej funkcie.

3. Pokúsime sa nahradiť informačnú maticu iným vyjadrením informácie, ktoré by nebolo závislé na θ . 4. Postavíme sa na stanovisko, že vystačíme s "lokálne optimálnymi" návrhmi experimentov. Poprípade preveríme lokálne najlepšie návrhy pre "rôzne lokality", a urobíme kompromis.

Žiaľ, postup uvedený v bode 3. nebýva použiteľný, pretože platí nasledujúce tvrdenie.

V PRÍPADE ODHADOVANIA PARAMETROV MNOŽSTVO INFORMÁCIE ZÍSKANEJ Z EXPERIMENTU PRINCIPIÁLNE ZÁVISÍ OD SKUTOČNEJ HODNOTY θ .

"Dôkaz" tohto tvrdenia vykonáme na obrázku. Napriek primitívnosti tohto postupu, myslím si, že presvedčí čitateľa.

Krivka na obrázku znázorňuje tzv. krivku stredných hodnôt, t.j. množinu

$$\{\eta(\theta) : \theta \in \Theta\}$$

v regresnom modeli $y = \eta(\theta) + \epsilon$ s jednorozmerným parametrom, $\dim(\theta) = 1$. Vzdialenosti medzi čiarkami vyznačenými na tejto krivke zodpovedajú konštantnému nárastu parametra θ pozdĺž krivky. Body $\eta(\theta^*)$, $\eta(\theta^\#)$, $\eta(\theta^\circ)$ zodpovedajú trom rôznym (hypotetickým) polohám skutočnej strednej hodnoty vektora y . Možné hodnoty pozorovaných vektorov y sú s veľkou pravdepodobnosťou v blízkosti svojej strednej hodnoty. Na obrázku sú vyznačené "mrakmi" bodov okolo

$$\eta(\theta^*), \eta(\theta^\#), \eta(\theta^\circ).$$

To isté platí aj pre odhady: tiež sú viac-menej koncentrované okolo θ^* , $\theta^\#$, θ° . Špeciálne, v prípade MNS odhady sa získavajú projekciou "mraku bodov" na krivku stredných hodnôt. Už z obrázku je jasné, že vzhľadom na rôzne zakrivenie krivky, a vzhľadom na nerovnomernosť parametrizácie tejto krivky, pri každej z hypotetických hodnôt θ^* , $\theta^\#$, θ° dostávame iný tvar rozdelenia pravdepodobnosti odhadu, inú koncentráciu odhadov okolo skutočného θ a teda aj rozdielne množstvo informácie z experimentu.

3.1 Vynútená linearizácia modelu

V lineárnych modeloch informačná matica nezávisí od θ a je dobrou mierou množstva informácie z experimentu. Situácia bude zrejme dosť dobrá, ak model je v nejakom zmysle takmer lineárny (napr. uvažujeme odhady nelineárnych parametrických funkcií v lineárnom modeli [10]). Je vcelku zrejmé, že pre rôzne návrhy experimentu sa dosahuje rôzny stupeň nelinearity modelu. (Např. dá sa ľahko ukázať, že v prípade, že návrh je koncentrovaný do rovnakého počtu bodov, ako je počet neznámych parametrov, tak regresný model zodpovedajúci tomuto návrhu je vnútorne lineárny). Pokiaľ by sme sa teda obmedzili len na také návrhy experimentu, ktoré dávajú modely blízke ku lineárnym, mohli by sme použiť kritéria optimality známe z lineárneho

modelu. Dá sa to vykonať napr. tak, že minimalizácia kritéria optimality sa vykonáva podmienene, za ohraňenej tzv. vnútornej krivosti a parametrickej krivosti modelu (pozri [12]). (Určité postupy v tomto smere sú napr. v práci [3].) Tým ovšem vylučujeme apriori tie návrhy, ktoré vedú k nelineárnym modelom, ale ktoré napriek tomu môžu byť informačne zaujímavé.

3.2 Sekvenčné navrhovanie nelineárných experimentov

Ak v experimente môžeme vykonať veľké množstvo pokusov, a ak technická stránka experimentu dovoľuje realizovať sekvenčné navrhovanie, toto navrhovanie experimentu dáva, zdá sa, dobré odpovede na otázky A) B) C) formulované v úvode. Sučasne však vznikajú nové problémy, ako hneď uvedieme.

Pri sekvenčnom navrhovaní najprv navrhujeme nesekvenčne určitú skupinu počiatočných pokusov

$$x_1, \dots, x_{N_1}$$

Každý ďalší pokus už navrhujeme sekvenčne, t.j. až po vykonaní všetkých predchádzajúcich pokusov, a po vypočítaní príslušných MNŠ odhadov zo súhrnu všetkých predchádzajúcich pokusov. Označme symbolom $\hat{\theta}^{(N)}$ odhad, ktorý takto získame po N pokusoch. Algoritmus pre výpočet ďalšieho pokusu x_{N+1} dostávame zvyčajne z algoritmov známych z navrhovania lineárných modelov. To znamená, že vypočítame informačnú maticu vzatú pre $\theta = \hat{\theta}^{(N)}$, t.j. maticu $M(x_1, \dots, x_N; \hat{\theta}^{(N)})$, a z tejto matice určíme algoritmus na N -tom kroku, ako keby to bola informačná matica lineárneho experimentu. O postupnosti takto vzniknutých bodov x_1, x_2, \dots množiny \mathcal{X} očakávame, že asymptoticky vytvorí návrhovú mieru, ktorá je v istom zmysle optimálna.

Na rozdiel od nesekvenčného navrhovania, vzniká tu nasledujúci rušivý jav. Body x_i sú vypočítané na základe predchádzajúcich pozorovaní a preto sú náhodné. Keďže vstupujú do stredných hodnôt ďalších pozorovaných veličín

$$y(x_{i+1}), y(x_{i+2}), \dots$$

menia rozdelenia pravdepodobnosti týchto veličín. Tak napríklad, zatiaľ čo pri nesekvenčnom navrhovaní sú veličiny $y(x_{i+1}), y(x_{i+2})$ nezávislé, pri sekvenčnom navrhovaní sa stávajú závislými. Tým ovšem strokotáva celá asymptotická teória odhadov v nelineárnom regresnom modeli. Je postupnosť takto sekvenčne vypočítaných odhadov ešte konzistentná, alebo asymptoticky normálna? Ak áno, je ich asymptotická variančná matica určená inverziou limitnej informačnej matice?

Riešením týchto obtiažnych teoretických otázok sa zaoberal rad autorov, medzi ktorými sú aj také známe mená ako Wu [19] alebo Woodroffe [18]. Dá sa povedať, že problém vyriešili za veľmi komplikovaných predpokladov. V jednom z posledných článkov [2] autori tvrdia, že našli jednoduchšie postačujúce podmienky, nato, aby dokázali nasledujúce tvrdenia.

1. Postupnosť normovaných informačných matíc vzatých v bodoch $\hat{\theta}^{(N)}$

:

$$\frac{1}{N}M(x_1, \dots, x_N; \hat{\theta}^{(N)}); \quad N = N_1, N_2, \dots$$

konverguje podľa pravdepodobnosti ku informačnej matici návrhovej miery ξ^*

$$M(\xi^*; \theta_{skut})$$

pričom ξ^* je lokálne D-optimálna návrhová miera pre lokalitu bodu $\theta = \theta_{skut}$, t.j. platí

$$\xi^* = \arg \max_{\xi} \det[M(\xi; \theta_{skut})]$$

Tu θ_{skut} označuje skutočnú hodnotu parametra θ .

2. Postupnosť náhodných vektorov

$$M^{1/2}(x_1, \dots, x_N; \hat{\theta}^{(N)})[\hat{\theta}_N - \theta]$$

konverguje podľa distribúcie ku normálnemu rozdeleniu $\mathcal{N}(0, I)$.

Z týchto dvoch tvrdení vyplýva aj pravidlo zastavenia sekvenčnej procedury, prinajmenšom heuristicky. Pre veľké N je $M(x_1, \dots, x_N, \theta^{(N)})$ blízke ku $M(x_1, \dots, x_N, \theta_{skut})$, experiment sa správa ako lineárny a pravidlo zastavenia bude presne také aké sa používa pri iteračných procedurách v lineárnych regresných experimentoch (pozri [11], veta IV.28).

3.3 Kritéria optimality založené na aproximáciách 2. rádu pre momenty a entropiu.

V ďalšom sa budeme zaoberať hľadaním kritérií optimality návrhu experimentu v tých prípadoch, keď rozsah návrhu je nedostatočný na to, aby sme MNS odhad mohli považovať za asymptoticky normálny.

Zakladná myšlienka aproximácii 2. alebo vyššieho rádu je nasledujúca: Odhad $\hat{\theta} = \hat{\theta}(y)$ aproximujeme pomocou Taylorovho rozvoja

$$\begin{aligned} \hat{\theta}(y) &= \hat{\theta}(y)|_{y=\eta(\theta_{skut})} + \frac{\partial \hat{\theta}(y)}{\partial y^T} \Big|_{y=\eta(\theta_{skut})} [y - \eta(\theta_{skut})] \\ &+ [y - \eta(\theta_{skut})]^T \frac{\partial^2 \hat{\theta}(y)}{\partial y \partial y^T} \Big|_{y=\eta(\theta_{skut})} [y - \eta(\theta_{skut})] + \dots \end{aligned}$$

Zrejme platí

$$\hat{\theta}(y)|_{y=\eta(\theta_{skut})} = \theta_{skut}$$

Tým je určený prvý člen rozvoja. Parciálne derivácie funkcie $\hat{\theta}(y)$, ktoré sú v ďalších členoch, vypočítame z normálnych rovníc pre MNS odhad, pretože tieto rovnice definujú odhad implicitne a jednoznačne v istom okolí bodu $y = \eta(\theta_{skut})$. Použijeme pri tom známu vetu o implicitnej funkcii.

Keď máme určené derivácie, ďalšie použitie Taylorovho rozvoja je jednoduché (aspoň v princípe). Napr., ak potrebujeme aproximácie momentov odhadu, stredujeme rozvoj resp. jeho mocniny člen po člene, a momenty odhadu určíme na základe známych momentov vektora $y - \eta(\theta_{skut})$. Podotýkame, že neustále predpokladáme, že tento vektor je normálne rozdelený.

Takýto postup vedie pomerne jednoducho ku aproximácii druhého rádu pre strednú hodnotu, avšak pri aproximácii 2. rádu pre variančnú maticu odhadu sa dostávame ku neprehľadným výrazom. Preto v práci [15] bola vypočítaná aproximácia 2. rádu pre entropiu rozdelenia pravdepodobnosti odhadu $\hat{\theta}$. Entropia je mierou variability odhadu a je teda vhodná ako kritérium optimality experimentu. Ak uvedenou metódou vypočítame aproximáciu entropie prvého rádu, dostaneme plnú zhodu s kritériom D-optimality. Ak použijeme aproximáciu 2. rádu dostaneme iné kritérium, a ako bolo ukázané na príklade v [15], dostávame aj iný optimálny návrh.

Pripomínáme, že aj tieto kritéria optimality sú "lokálne", t.j. závislé od skutočnej hodnoty parametra θ .

3.4 Integrálne kritéria optimality

Pretože všetka informácia o vlastnostiach odhadu parametrov je v jeho distribúcii, uvedieme tu kritéria optimality, ktoré vychádzajú z tejto distribúcie (presnejšie, z hustoty pravdepodobnosti odhadu).

Aproximácie momentov alebo entropie uvedené v predchádzajúcom paragrafe môžu byť totiž nevhodné ak napr. rozdelenie pravdepodobnosti odhadu je bimodálne, alebo ak s nezanedbatelnou pravdepodobnosťou odhad leží na hranici parametrického priestoru Θ . Ale aj v menej extrémnych prípadoch aproximácie 2. a vyššieho rádu môžu byť dosť nepresné, jednoducho preto, že sú lokálne t.j. vychádzajú z Taylorovej formule lokalizovanej v jedinom bode (v bode θ_{skut}).

Omnoho presnejšie aproximácie možno docieľiť ak sa použije aproximácia hustoty pravdepodobnosti uvedená v [12]. Ak napr. za kritérium optimality zvolíme súčet stredne kvadratických odchýlok odhadov $\hat{\theta}_i$, toto kritérium je vyjadrené v tvare integrálu

$$\int_{\Theta} \|\hat{\theta} - \theta_{skut}\|^2 q(\hat{\theta}|\theta_{skut})$$

kde $q(\hat{\theta}|\theta_{skut})$ je hustota pravdepodobnosti odhadu, resp. jej aproximácia. Keďže rozdelenie pravdepodobnosti odhadu nemá hustotu na hranici, takýto

integrál by nemusel vyjadrovať príslušné kritérium optimality. Aby sa tento nedostatok odstránil, bol v [14] zvolený nasledujúci postup. Namiesto MNS odhadu $\hat{\theta}$ sa použil penalizovaný MNS odhad

$$\tilde{\theta}(y) = \arg \min_{\theta} \{ \|y - \eta(\theta)\|^2 + \omega(\theta) \}$$

Tu penalizácia $\omega(\theta)$ je nejaká nezáporná funkcia, dvakrát spojitely diferencovateľná na vnútre množiny Θ , rastúca ak θ sa blíži ku hranici množiny Θ , rovnajúca sa $+\infty$ ak θ je na hranici množiny Θ a nulová na nejakej vnútornej časti Θ . Je zrejmé, že odhad $\tilde{\theta}$ nikdy nedosiahne hranicu množiny Θ . Pomocou penalizačnej funkcie sa tie odhady $\hat{\theta}$, ktoré pôvodne ležali na hranici Θ , posunú smerom dovnútra Θ . Pritom, ak je penalizácia $\omega(\theta)$ vhodne zvolená, integrál sa zámennou odhadu $\tilde{\theta}$ za odhad $\hat{\theta}$ priliš nezmení.

Použitá hustota pravdepodobnosti odhadu $\hat{\theta}$ má tvar

$$q(\hat{\theta}|\theta) = \frac{\det[Q(\hat{\theta}, \theta)]}{(2\pi)^{p/2} \det^{1/2}[M(\hat{\theta})]} \exp\left\{-\frac{1}{2} \|P(\hat{\theta})[\eta(\hat{\theta}) - \eta(\theta)]\|^2\right\}$$

kde

$$Q_{ij}(\hat{\theta}, \theta) = M_{ij}(\hat{\theta}) + [\eta(\hat{\theta}) - \eta(\theta)]^T [I - P(\hat{\theta})] \frac{\partial^2 \eta(\hat{\theta})}{\partial \hat{\theta}_i \partial \hat{\theta}_j}$$

a kde

$$P(\theta) = \frac{\partial \eta(\theta)}{\partial \theta^T} M^{-1}(\theta) \frac{\partial \eta^T(\theta)}{\partial \theta}$$

V prípade hustoty odhadu $\tilde{\theta}$ dochádza v predchádzajúcom vzorci pre hustotu k nasledujúcim zmenám. Vektor $\hat{\theta}$ nahradíme vektorom $\tilde{\theta}$. Ku matici $Q_{ij}(\tilde{\theta}, \theta)$, ktorej determinant je uvedený v čitateli, pripočítame maticu

$$\frac{\partial^2 \omega(\tilde{\theta})}{\partial \tilde{\theta}_i \partial \tilde{\theta}_j} - \frac{1}{\sigma^2} u^T(\tilde{\theta}) \frac{\partial^2 \eta(\tilde{\theta})}{\partial \tilde{\theta}_i \partial \tilde{\theta}_j}$$

a v exponente pripočítame ku vektoru $\eta(\tilde{\theta})$ vektor

$$u(\tilde{\theta}) = \frac{\partial \eta(\tilde{\theta})}{\partial \tilde{\theta}^T} M^{-1}(\tilde{\theta}) \frac{\partial \omega(\tilde{\theta})}{\partial \tilde{\theta}}$$

Aj keď tieto hustoty pravdepodobnosti sú dané explicitnými vzorcami a možno ich ľahko numericky spočítať, integrál vyjadrujúci kritérium optimality nevieme odstrániť. Pri výpočte optimálneho návrhu musíme teda minimalizovať integrál. Tu sa ukazuje výhodným použitie metód stochastickej aproximácie, tak ako je to uvedené v [14] alebo v [15].

3.5 Kritéria optimality založené na aproximácii objemu oblasti spoľahlivosti pre θ

V prácach [7], [17] je navrhnuté aproximovať objem oblasti spoľahlivosti pre θ . Takéto aproximácie sú však veľmi diskutabilné, a sú kritizované aj v [5]. Príčin je niekoľko: a) V nelineárnom modeli neexistuje jednotná koncepcia oblasti spoľahlivosti. Tak aj autori uvedených prác používajú rôzne oblasti. b) Aproximácie objemu sú lokálne, opäť pomocou nejakého rozvoja, a teda omnoho nepresnejšie ako aproximácie hustôt pravdepodobnosti uvedené vyššie. c) Oblasť spoľahlivosti a aj jej objem sú náhodne, ich predikcia vyžaduje ďalšie stredovanie. Podľa môjho názoru sú kritéria optimality založené na mierach koncentrácie hustoty pravdepodobnosti odhadu omnoho prijateľnejšie než kritéria optimality založené na koncentrácii "konfidenčnej miery".

3.6 Bayesovské kritéria optimality

Bayesovské kritéria optimality vyžadujú znalosť dvoch vecí

- apriórneho rozdelenia pravdepodobnosti na Θ daného napr. hustotou $\pi(\theta)$.
- užítkovej funkcie

$$U[d, \theta, x^*, y]$$

závislej vo všeobecnosti od rozhodnutia d , od parametra θ , od návrhu experimentu $x^* := x_1, \dots, x_N$, ako aj od vektora nameraných údajov y .

V duchu bayesovskej štatistiky najprv zvolíme rozhodovaciu funkciu $\Delta(y)$, tak, aby sme maximalizovali aposteriorný stredný užitek:

$$\Delta(y) = \arg \max_d \int_{\Theta} U[d, \theta, x^*, y] p(\theta|y, x^*) d\theta$$

kde $p(\theta|y, x^*)$ je aposteriorna hustota pravdepodobnosti. Zodpovedajúcim kritériom optimality je celkový stredný užitek a tento je vyjadrený ako funkcia návrhu x^* tvaru

$$\Phi(x^*) = \int_{R^N} \left[\int_{\Theta} U[\Delta(y), \theta, x^*, y] p(\theta|y, x^*) d\theta \right] p(y|x^*) dy$$

kde $p(y|x^*)$ je prediktívna hustota pravdepodobnosti vektora y . Toto kritérium optimality nezávisí od skutočnej hodnoty vektora parametrov θ a teda otázky A) a B) z úvodu sú tu zodpovedané, hlavne vďaka znalosti apriórneho rozdelenia, ktoré dovoľuje rôzne predikcie, včítane predikcie informácie získanej z experimentu. Otázka výpočtu optimálneho návrhu je diskutovaná v prehľadovom článku [1]

3.7 Použitie návrhovej miery v nelineárnom modeli

V princípe nie je obtiažne nahradiť návrh x_1, \dots, x_N zodpovedajúcou návrhovou mierou ξ , presne tak ako v lineárnom modeli. Na to stačí, aby vo všetkých výrazoch, kde sa vyskytujú sumy podľa x_1, \dots, x_N , tieto boli nahradené integrálmi podľa miery ξ . Ide o jednoduchú úpravu. Jej výhodou je, že diskrétny problém určenia návrhu x_1, \dots, x_N možno takto nahradiť spojitým problémom určenia optimálnej miery ξ . Kritéria optimality, ktoré vystupujú v nelineárnych problémoch, nie sú však konvexnými funkciami návrhovej miery, teda užitočnosť použitia návrhovej miery je dosť diskutabilná.

References

- [1] K. Chaloner and I. Verdinelli: Bayesian experimental design: a review. *Statist. Science* **10** (1995), 273-304.
- [2] P. Chaudhuri and P.A. Mykland: Nonlinear experiments: optimal design and inference based on likelihood. *J. Amer. Stat. Assoc.* **88** (1993), 538-546.
- [3] M.A. Clyde: Bayesian designs for approximate normality. In: *Advances in Model Oriented Data Analysis*, p. 25-36. Eds. Ch.P. Kitsos and W.G. Müller. Physica-Verlag, Heidelberg, 1995.
- [4] V.V. Fedorov and Ch. Nachtsheim: Optimal design for time dependent responses. In: *Advances in Model Oriented Data Analysis*, p. 3-14. Eds. Ch.P. Kitsos and W.G. Müller. Physica-Verlag, Heidelberg, 1995.
- [5] I. Ford, C.P. Kitsos and D.M. Titterton: Recent advances in nonlinear experimental design. *Technometrics* **31** (1985), 49-60.
- [6] N. Gaffke and B. Heiligers: Algorithms for optimal design with application to multiple polynomial regression. *Metrika* **42** (1995) 173-190.
- [7] D.C. Hamilton: A quadratic design criterion for precise estimation in nonlinear regression models. *Technometrics* **27** (1985), 241-250.
- [8] J. Kiefer and J. Wolfowitz: Optimum designs in regression problems. *Ann. Math. Stat.* **30** (1959), 271-194.
- [9] J. Mikulecká: Hybrid experimental design. *Kybernetika* **19** (1983), 1-14.
- [10] Ch. Müller: Maximin efficient designs for estimating nonlinear aspects in linear models. *J. Statist. Planning and Inference.* **44** (1995), 117-132.

- [11] A. Pázman: Základy optimalizácie experimentu. Veda, Bratislava, 1980.
- [12] A. Pázman: Nonlinear Statistical Models. Dordrecht, Kluwer Acad. Publ. 1993.
- [13] A. Pázman: Distribution of estimators and optimal experimental design in nonlinear regression. Kybernetika (1996) - v tlači
- [14] A.Pázman and L.Pronzato: Nonlinear experimental design based on the distribution of estimators. J.Statist. Planning anf Inference. **33** (1992), 385-402.
- [15] L. Pronzato and A. Pázman: Second-order approximation of the entropy in nonlinear least-squares estimation. Kybernetika **30** (1994), 187-198. Erratum. Kybernetika **32** (1996), 104.
- [16] F. Pukelsheim: Optimal Design of Experiments. Wiley , N.Y. 1993.
- [17] J.P.Vila: Optimal designs for parameter estimations in nonlinear regression models. Proc. of the XIIIth Inter. Biometric Conference, Seatle, 1986.
- [18] M. Woodroofe: Very weak expansions for sequentially designed experiments. The Annals of Statistics **17** (1989), 1087-1102.
- [19] C.F.J.Wu: Asymptotic inference from sequential design in nonlinear situation. Biometrika **72** (1985), 553-558.