

1. ÚVOD

S robustními statistickými metodami se dnes setkáváme v mnoha oblastech matematické statistiky. Existuje značné množství časopiseckých článků i knižních publikací. Cílem tohoto příspěvku je uvést některé výsledky týkající se aproximace jednoho z robustních odhadů, studentizovaného M-odhadu, pomocí tzv. k-krokové verze, a to pro případ absolutně spojitě i schodovitě generující funkce.

M-odhady parametru polohy, zavedené původně Huberem, jsou invariantní vzhledem k posunutí, ale nejsou ekvivariantní vzhledem ke změně měřítka, což v praxi přináší určité problémy. Ty mohou být odstraněny např. tím, že budou společně odhadovány parametry posunutí a měřítka - zde je ale problematické interpretovat odhadované parametry. Jinou možností, uvažovanou i v tomto textu, je použití studentizovaných M-odhadů.

2. ZÁKLADNÍ SITUACE

Nechť X_1, X_2, \dots je posloupnost nezávislých náhodných veličin se společnou absolutně spojitou distribuční funkcí $F(x-\theta)$, která není obecně známa. Cílem je konstrukce takového odhadu parametru polohy θ , jehož definice závisí pouze na X_1, X_2, \dots, X_n a je nezávislá na tvaru funkce F . Takový je například klasický odhad metodou nejmenších čtverců, ale různé teoretické i numerické studie ukazují jeho vysokou citlivost na porušení předpokladu normality, na výskyt odlehlých pozorování. Proto je vhodné použití robustních metod.

Definice 1.

Nechť $\varphi(x)$ je spojitá funkce $\mathbb{R}^1 \rightarrow \mathbb{R}^1$. Nechť $S_n = S_n(X_1, X_2, \dots, X_n)$ je statistika, které splňuje

$$(1) \quad \begin{aligned} S_n(x) &> 0 && \text{s.j.} \\ S_n(x+c) &= S_n(x) \\ S_n(cx) &= c S_n(x) \end{aligned} \quad ; c > 0, x \in \mathbb{R}^n$$

a nechť existuje funkcionál $S = S(F) > 0$ takový, že

$$(2) \quad n^{1/2} (S_n - S(F)) = O_p(1) \quad \text{při } n \rightarrow \infty.$$

Potom studentizovaným M-odhadem parametru θ generovaným funkcí $\varphi(x)$ nazveme statistiku $M_n = M_n(X_1, X_2, \dots, X_n)$, která je řešením minimalizace

$$(3) \quad \sum_{i=1}^n \varphi((X_i - t) / S_n) := \min \quad \text{vzhledem k } t \in \mathbb{R}^1.$$

Jestliže $\psi = \varphi'$ je absolutně spojitá funkce, je M_n jedním z řešení rovnice

$$(4) \quad \sum_{i=1}^n \psi((X_i - t) / S_n) = 0.$$

Je-li φ konvexní funkce a tedy ψ neklesající, je M_n určen jednoznačně ve tvaru

$$(5) \quad M_n = \frac{1}{2} (M_n^+ + M_n^-), \quad \text{kde}$$

$$(6) \quad \begin{aligned} M_n^- &= \sup \left\{ t: \sum_{i=1}^n \psi((X_i - t) / S_n) > 0 \right\} \quad \text{a} \\ M_n^+ &= \inf \left\{ t: \sum_{i=1}^n \psi((X_i - t) / S_n) < 0 \right\}. \end{aligned}$$

V případě, že ψ je spojitá, ale φ obecně není konvexní, má rovnice (4) obecně více řešení. Za určitých obecných podmínek existuje alespoň jedno řešení (4), které je $n^{1/2}$ -konzistentním odhadem θ , tj. které při $n \rightarrow \infty$ splňuje

$$(7) \quad n^{1/2} (M_n - \theta) = O_p(1).$$

Jestliže nějakým výpočetním postupem (většinou iteračním) dospějeme k nějakému kořenu rovnice (4), ještě nevíme, je-li toto řešení $n^{1/2}$ -konzistentním odhadem Θ . V této situaci můžeme zvolit jeden z následujících postupů:

a/ Ze všech kořenů rovnice (4) volíme ten, který je nejbližší danému počátečnímu $n^{1/2}$ -konzistentnímu odhadu Θ , např. výběrovému mediánu.

b/ Vyjdeme z počátečního $n^{1/2}$ -konzistentního odhadu a vypočteme jednokrokovou verzi M-odhadu, jak bude ukázáno níže.

V případě, kdy φ je absolutně spojitá funkce, ale její derivace ψ má body nespojitosti 1. druhu, nemusí mít rovnice (4) s kladnou pravděpodobností vůbec řešení. Mírné předpoklady zejména o konvexitě funkce φ zajistí, že M-odhad M_n je i v tomto případě určen jednoznačně vztahy (5) a (6). Tento odhad je zároveň $n^{1/2}$ -konzistentní, pokud předpokládáme existenci dvou ohraničených derivací funkce F v okolí bodů nespojitosti funkce ψ .

V případě, že φ není konvexní a ψ je funkce se skoky, neumíme dokázat našimi metodami $n^{1/2}$ -konzistenci řešení minimalizace (3).

3. VOLBY FUNKCE

V [1] jsou popsány různé volby funkcí φ a ψ . Nejčastěji se pracuje s Huberovou funkcí

$$(8) \quad \psi(x) = \begin{cases} x & \text{pro } |x| \leq d \\ d \operatorname{sign} x & \text{pro } |x| > d, \quad d > 0. \end{cases}$$

Nejjednodušší ze schodovitých funkcí ψ je funkce

$$(9) \quad \psi(x) = \begin{cases} 0 & \text{pro } x \leq r \\ 1 & \text{pro } x > r, \quad r \in \mathbb{R}^1. \end{cases}$$

Libovolnou schodovitou funkcí s konečným počtem skoků lze pak vyjádřit pomocí nějaké lineární kombinace funkcí (9), které mají jediný skok o velikosti 1.

4. JEDNOKROKOVÁ VERZE STUDENTIZOVANÝCH M-ODHADŮ

Odhady definované implicitně (např. určené jako řešení nelineární rovnice) se často hledají pomocí nějakých aproximací, jejichž výpočet je snadnější a u nichž lze určit, jak přesné přiblížení k odhadu poskytují. K takovým aproximacím patří i jednokroková verze M-odhadu. Poprvé ji v souvislosti s M-odhadem uvedl Bickel (1975, [2]), i když základní myšlenka výpočtu právě jedné Newtonovy iterace je podstatně starší.

Ukazuje se, že jednokrokovou verzi M-odhadu asymptoticky ekvivalentní M_n lze odvodit za určitých podmínek regularity na funkce ψ a F ; hlavně však musíme předpokládat existenci vhodného $n^{1/2}$ -konzistentního počátečního odhadu $M_n^{(0)}$ parametru Θ . Předpokládejme proto, že existuje $M_n^{(0)}$ tak, že

$$(10) \quad n^{1/2} (M_n^{(0)} - \Theta) = o_p(1), \quad \text{když } n \rightarrow \infty.$$

Definice 2.

Jednokrokovou verzi studentizovaného M-odhadu nazveme veličinu

$$(11) \quad M_n^{(1)} = \begin{cases} M_n^{(0)} + (n \hat{g}^{(1)})^{-1} \sum_{i=1}^n \psi((X_i - M_n^{(0)})/S_n) & \text{pro } \hat{g}^{(1)} \neq 0 \\ M_n^{(0)} & \hat{g}^{(1)} = 0, \end{cases}$$

kde

$$(12) \quad \hat{g}^{(1)} = n^{-1/2} (t_2 - t_1)^{-1} \sum_{i=1}^n [\psi((X_i - M_n^{(0)} + n^{-1/2} t_2) / S_n) - \psi((X_i - M_n^{(0)} + n^{-1/2} t_1) / S_n)], \quad t_1 < t_2,$$

je konzistentní odhad funkcionálu $g = g(\psi, F)$, který je definován jako

$$(13) \quad g = S^{-1} \int \psi'(x/S) dF(x) \quad \text{v případě absolutně spojitě funkce } \psi \text{ a jako}$$

$$(14) \quad g = \sum_{j=1}^q (a_j - a_{j-1}) f(Sr_j) \quad \text{v případě schodovité funkce } \psi, \text{ přičemž}$$

f je derivace distribuční funkce F ; r_1, r_2, \dots, r_q jsou body nespojitosti funkce ψ a a_0, a_1, \dots, a_q jsou daná reálná čísla.

Asymptotické vlastnosti jednokrokových verzí M -odhadů a přesnost aproximace M_n pomocí $M_n^{(1)}$ lze studovat za pomoci vět o asymptotické linearitě a reprezentaci (viz [4]). Za určitých dalších předpokladů (zejména na hladkost funkce ψ) umožňuje asymptotická linearita stanovit přesný řád konzistence $\hat{\gamma}^{(1)}$ a tudíž i přesný řád $M_n^{(1)} - M_n$.

Pro spojitou funkci ψ se odvodí

$$(15) \quad M_n^{(1)} - M_n = O_p(n^{-1})$$

a pro ψ schodovitou

$$(16) \quad M_n^{(1)} - M_n = O_p(n^{-3/4}).$$

Ukazuje se, že řád asymptotické ekvivalence i asymptotické rozdělení jednokrokové verze jsou stejné pro všechny volby $n^{1/2}$ -konzistentního počátečního odhadu. Závislost chování jednokrokové verze odhadu na volbě $M_n^{(0)}$ se projevuje až ve členech druhého řádu (viz [3]).

5. K-KROKOVÁ VERZE STUDENTIZOVANÝCH M -ODHADŮ

Dosadíme-li v (11) a (12) místo $M_n^{(0)}$ jednokrokovou verzi $M_n^{(1)}$, dostaneme 2. iteraci, která obecně lépe aproximuje M_n než $M_n^{(1)}$ (dvoukroková verze $M_n^{(2)}$). Tato situace, ale bez studentizace, je uvažována v [5]. Rekurentně takto můžeme vytvořit k -krokovou verzi pro libovolné kladné celé k , rekurentně se samozřejmě počítá i odhad $\hat{\gamma}^{(k)}$ funkcionálu γ .

Přesnost aproximace M -odhadu k -krokovou verzí lze nyní opět stanovit pomocí vět o asymptotické linearitě a reprezentaci, které je však nutno podstatným způsobem modifikovat. V případě spojitě funkce ψ nyní platí

$$(17) \quad n^{1/2}(\hat{\gamma}^{(k)} - \gamma) = O_p(1)$$

$$(18) \quad n^{1/2}(M_n^{(k)} - M_n) = O_p(n^{-k/2}), \quad k=1,2,\dots, n \rightarrow \infty.$$

V tomto případě mají tedy vícekové verze všechny vhodné vlastnosti jednokrokových verzí (snadný výpočet, konvergence ke konzistentnímu řešení) a navíc i lepší řád rozdílu $M_n^{(k)} - M_n$. V případě schodovité funkce ψ se však ukazuje, že řád konvergence rozdílu $M_n^{(k)} - M_n$ k nule se s rostoucím k mění podstatně méně než pro hladkou funkci ψ a nikdy není lepší než $O_p(n^{-1})$. V tomto případě zřejmě pro praktické účely dobře poslouží jednokroková verze a výpočet dalších verzí se jeví jako neefektivní.

6. LITERATURA

- [1] Andrews, D.F. (1972). Robust Estimates of Location. Princeton University Press, New Jersey.
- [2] Bickel, P.J. (1975). One-step Huber estimates in the linear model. J.Amer.Statist. Assoc. 70, 428-433.
- [3] Jurečková, J. (1985). Robust estimation of location and their second order asymptotic relations. A Celebration of Statistics A.C. Atkinson and S.A. Fienberg, ed., 377-392. Springer, New York.
- [4] Jurečková, J. (1986). Asymptotic representation of L -estimators and their relations to M -estimators. Sequential Analysis 5, 317-338.
- [5] Janssen J., Jurečková J., Veraverbeke N. (1975). Rate of convergence of one- and two-step M -estimators with applications to maximum likelihood and Pitman estimators. Ann.Statist. 13, 1222-1229.
- [6] Welsh, A.H. (1989). On M -processes and M -estimation. Ann.Statist. 17, 337-361.

Jiří Militký, KPM, VŠST Liberec

1. Úvod

Pro třídu personálních počítačů je v současné době k dispozici celá řada programových paketů umožňujících řešení úlohy nelineární regrese. Řada z nich však obsahuje dosti výrazné nedostatky jak z hlediska numerického (hledání stacionárního bodu), tak i statistického (analýza odhadů, modelu a dat). Důsledkem je, že řada uživatelů vlastně a priori nemůže svoji úlohu s využitím takového software korektně vyřešit. Jedním ze základních problémů je konvergence k falešným extrémům, kdy je sice formálně řešení nalezeno, ale nespĺňuje podmínky extrému kritéria regrese.

Poměrně nejrozvinutější stránkou programových profesionálních paketů je kvalita okolí (ovládání vstupů / výstupů, řízení programů atd.). Trendy v této oblasti obsahuje práce /1/.

Cílem tohoto příspěvku je stručně diskutovat pouze numerické a statistické zvláštnosti nelineární regrese a porovnat z těchto hledisek nejznámější statistické pakety. S ohledem na to, že máme k dispozici také vlastní statistický paket využívající pro regresi původní program MINOPT, je provedeno jeho zařazení mezi testované pakety. Provedená testovací studie je sice rozsahem nevelká, ale byly úmyslně vybrány speciální příklady, které činí při použití standardních algoritmů obtíže.

2. Základní pojmy

Mějme k dispozici n -tici naměřených dat $\{y_i, x_i\}$ $i=1 \dots n$, o nichž předpokládáme, že odpovídají aditivnímu modelu působení chyb, t.j.

$$y_i = f(x_i, \underline{a}) + \varepsilon_i \quad (1)$$

V rov. (1) je $f(x_i, \underline{a})$ funkce nelineární alespoň vzhledem k jedné složce vektoru modelových parametrů $\underline{a} = (a_1, \dots, a_m)^T$. Úlohou nelineární regrese je pak získat bodové odhady $\hat{\underline{a}}$ těchto parametrů a jejich statistické charakteristiky.

V řadě případů lze chyby ε_i považovat za nezávislé, stejně rozdělené náhodné veličiny s nulovou střední hodnotou $E(\varepsilon_i) = 0$ a konstantním rozptylem $D(\varepsilon_i) = \sigma^2$. Odhady $\hat{\underline{a}}$ lze za těchto podmínek určit minimalizací kritéria nejmenších čtverců

$$S(\underline{a}) = \sum_{i=1}^n (y_i - f(x_i, \underline{a}))^2 \quad (2)$$

Pokud chyby ε_i pocházejí z normálního rozdělení $N(0, \sigma^2)$, jsou odhady $\hat{\underline{a}}$ minimalizující kritérium $S(\underline{a})$ maximálně věrohodné a platí, že jsou asymptoticky normální

$$\hat{\underline{a}} \rightarrow N(\underline{a}, \sigma^2 (\mathcal{J}^T \mathcal{J})^{-1}) \quad (3)$$

Asymptotická normalita nevyžaduje obecně normalitu chyb, ale $f(x, \underline{a})$ musí být spojitá v prvních dvou derivacích, matice $\mathcal{J}^T \mathcal{J}$ musí být regulární a úloha musí být identifikovatelná /4/. V zápisu (3) je symbolem \mathcal{J} označena matice prvních derivací modelové funkce podle jednotlivých parametrů s prvky

$$J_{ij} = \frac{\partial f(x_i, \underline{a})}{\partial a_j} \quad (4)$$

$$j = 1, \dots, m$$

$$i = 1, \dots, n$$

Programy pro nelineární regresi musí obecně nejdříve nalézt minimum funkce $S(\underline{a})$ a pak následně provést statistickou analýzu regrese (s ohledem na parametry, data a předpoklady vedoucí ke kritériu nejmenších čtverců).

V některých úlohách lze specifikovat jiné rozdělení chyb, nebo uvažovat porušení některých základních předpokladů (obvykle nekonstantnost rozptylu, nebo závislosti jednotlivých chyb). Pak se buď místo klasické metody nejmenších čtverců (MNČ) používá vážená (VMNČ), nebo se volí speciální kritéria. Pro třídu symetrických hustot lišících se od normálního rozdělení jen délkou konců se často aplikuje tzv. L_p - kritérium regrese

$$L_p(\underline{a}) = \sum_{i=1}^n (y_i - f(x_i, \underline{a}))^p \quad (5)$$

kde $1 \leq p < \infty$ je parametr související s délkou konců rozdělení chyb. Zde se v dalším omezíme pouze na kritérium MNČ s tím, že v části o jednotlivých programových paketech uvedeme, které možnosti navíc lze použít.

3. Numerická stránka nelineární regrese

Naprostá většina programů pro minimalizaci kritéria $S(\underline{a})$ je iterační povahy. Obecně lze pro řešení této úlohy použít derivačních, resp. nederivačních algoritmů /3/. Zde se omezíme pouze na derivační využívající k činnosti speciální tvar kritéria $S(\underline{a})$. Celý postup hledání minima se pak skládá z těchto kroků :

1. Stanovení výchozího (počátečního) odhadu parametrů $\underline{a}^{(0)}$ pro první iteraci, resp. startovacího odhadu $\underline{a}^{(i)}$ pro i -tou iteraci (obvykle jde o výsledek $(i-1)$ ní iterace).
2. Nalezení přijatelného směrového vektoru \underline{v}_i .
3. Určení vhodného přírůstkového vektoru $\underline{\Delta}_i$. To je obvykle ekvivalentní výpočtu skaláru α_i tak, aby

$$S(\underline{a}^{(i)} + \underline{\Delta}_i) \leq S(\underline{a}^{(i)}) \quad (6)$$

4. Ověření toho, zda výsledek i -té iterace $\underline{a}^{(i+1)} = \underline{a}^{(i)} + \underline{\Delta}_i$ neodpovídá stacionárnímu bodu $\hat{\underline{a}}$.

Tento cyklus se iterativně opakuje. Jednotlivé algoritmy nelineární regrese se liší zejména ve způsobu nalezení přijatelného směrového vektoru \underline{v}_i a skaláru α_i . O zdaru minimalizace však spolurozhodují také způsoby realizace dalších fází uvedeného iteračního cyklu.

Pro stanovení prvních odhadů $\underline{a}^{(0)}$ pro speciální typy funkcí existuje v literatuře řada technik. Žádná však není zcela obecná, a proto všechny programové balíky pro nelineární regresi vyžadují zadání $\underline{a}^{(0)}$ od uživatele. Pro ukončení iteračního procesu se běžně používá pouze sledování relativních změn velikosti parametrů v po sobě následujících iteracích resp. relativních změn velikosti $S(\underline{a}^{(i)})$. Minimální přípustné změny navíc většinou zadává uživatel. To vede často ke stavu (zejména u silně nelineárních modelů), kdy je proces hledání minima ukončen v bodě značně vzdáleném od stacionárního. Řada programů nijak netestuje, zda nalezený vektor $\hat{\underline{a}}$ je či není stacionárním bodem, což uživateli neumožňuje rozhodnutí, zda je z numerického hlediska vše v pořádku.

Jednoduše lze ověřit kvalitu odhadů z hlediska numerického sledování velikosti gradientu \underline{g}_i kritériální podmínky $S(\underline{a})$. Pokud např. platí, že

$$\| \underline{g}_i \|_2^2 < \epsilon \quad (7)$$

považuje se \underline{a}^1 za extrém. Pro gradient g_i platí jednoduchý vztah

$$g_i = -2 J^T \underline{e} \quad (8)$$

kde \underline{e} je vektor reziduí s prvky

$$e_j = y_j - f(x_j, \underline{a}^{(i)}) \quad j=1, \dots, n \quad (9)$$

Velikost ϵ se volí obyčejně kolem 10^{-4} .

Další jednoduchá technika vychází z požadavku, že v minimu $S(\underline{a})$ je vektor reziduí \underline{e} kolmý na sloupce matice J . Pro kosinus úhlu β_j mezi \underline{e} a j -tým sloupcem J_j matice J zřejmě platí

$$\cos(\beta_j) = \underline{e}^T J_j [\|J_j\| \cdot \|\underline{e}\|]^{-1} \quad j=1, \dots, n \quad (10)$$

Pokud je maximální hodnota $\cos(\beta_j)$ dostatečně malá (např. menší než 10^{-9}), předpokládá se, že bylo nalezeno minimum.

Při určování přírůstkového vektoru Δ_i se využívá Taylorova rozvoje $S(\underline{a})$ do kvadratických členů včetně

$$S(\underline{a}) \approx S(\underline{a}^{(i)}) + (\underline{a} - \underline{a}^{(i)})^T g_i + \frac{1}{2} (\underline{a} - \underline{a}^{(i)})^T H_i (\underline{a} - \underline{a}^{(i)}) \quad (11)$$

Zde H_i je matice druhých derivací kritériální podmínky, kterou lze zapsat ve tvaru

$$H_i = 2 [J^T J + B] \quad (12)$$

Matice B má prvky

$$B_{lj} = \sum_{k=1}^n a_k \frac{\partial f(x_k, \underline{a}^{(i)})}{\partial a_l \partial a_j} \quad l, j = 1, \dots, m \quad (13)$$

Pro směrovou derivaci kritériální podmínky ve směru \underline{s} platí

$$\frac{\partial S(\underline{a})}{\partial \underline{a}} \Big|_{\underline{s}} = g^T \underline{s} = -2 J^T \underline{e} \underline{s} \quad (14)$$

Protože je účelem hledat minimum $S(\underline{a})$, musí být směrová derivace záporná, t.j. $g^T \underline{s} < 0$. Na základě této podmínky lze snadno určit, že přijatelný směrový vektor \underline{v}_i má tvar

$$\underline{v}_i = -R g \quad (15)$$

kde R je regulární pozitivně definitní matice. Pro nalezení optimálního skaláru α_i pro směr \underline{v}_i je možné využití rov. (11) při náhradě vektoru $(\underline{a} - \underline{a}^{(i)})$ vektorem $\alpha_i \underline{v}_i$. Po analytické minimalizaci vzhledem k α_i rezultuje vztah

$$\alpha_i^* = g^T R g [g^T R^T H R g]^{-1} \quad (16)$$

Rov. (16) definuje tzv. Raleighův koeficient a volba α_i^* zajišťuje nalezení lokálního minima $S(\underline{a})$ ve směru \underline{v}_i (za předpokladu, že lze kritériální podmínku dobře aproximovat jejím kvadratickým rozvojem).

Z rov. (15) je zřejmé, že v závislosti na volbě matice R existuje řada přijatelných směrových vektorů. Lze snadno určit i lokálně optimální směr, který snižuje $S(\underline{a})$ nejvíce. Ne vždy je však tento postup úspěšný, protože lokální optimalita v úlohách nelineární regrese běžně neznamená globální optimalitu.

Na základě Taylorova rozvoje $S(\underline{a})$ do kvadratických členů včetně ve směru \underline{v} a analytické minimalizace lze nalézt lokálně optimální přírůstek N_j ve tvaru

$$N_j = -H^{-1} g = (J^T J + B)^{-1} J^T \underline{e} \quad (17)$$

Směr N_j odpovídá Newtonově metodě a platí pro něj, že $\alpha_i^* = 1$. Pro případy, kdy je $S(\underline{a})$ kvadratická funkce, vede použití Newtonovy metody k optimu

v jednom kroku. V místech vzdálenějších od minima (kde již kvadratická aproximace $S(\underline{a})$ nepostačuje) však tato metoda příliš nevyhovuje. Navíc vyžaduje zadání (nebo přibližný výpočet) druhých derivací modelové funkce. Existuje celé spektrum metod variabilní metriky, (kvazimewtonovských), kde se provádí zpřesňování matice H resp. H^{-1} na základě informací o gradientech g v jednotlivých iteracích.

V blízkosti optima jsou složky vektoru \underline{e} blízké nule, takže lze matici B považovat za přibližně nulovou. Odpovídající směrový vektor má pak tvar

$$\underline{L}_j = (J^T J)^{-1} J^T \underline{e} \quad (18)$$

Lze dokázat, že tento směr odpovídá přírůstku určenému z linearizace modelové funkce $f(x, \underline{a})$. Metody využívající směrového vektoru \underline{L}_j se označují jako Gauss-Newtonovy. Pokud je $H \approx 2J^T J$, rezultuje $\alpha_j^* = 1$. V oblastech vzdálenějších od minima to však neplatí a je výhodnější volit zkrácenou délku kroku např. dle vztahu

$$\alpha_{1j}^* = 0.2 + 0.8 \|g_j\|^4 [g_j^T (J^T J)^{-1} g_j \cdot g_j^T (J^T J) g_j]^{-1} \quad (19)$$

Gauss - Newtonovy metody jsou v praxi došti oblíbené pro svoji poměrně rychlou konvergenci v jednodušších (nevýrazně nelineárních) úlohách. Pokud nelze dobře aproximovat $S(\underline{a})$ kvadratickým rozvojem, je možné volit jako lokálně optimální přírůstkový vektor ve směru záporně vzatého gradientu

$$\underline{A}_j = -g_j \quad (20)$$

Protože je zřejmé $R = E$, vyjde po dosazení do rov. (16)

$$\alpha_{2j}^* = g_j^T g_j [g_j^T (J^T J) g_j]^{-1} \quad (21)$$

Metody využívající směrového vektoru \underline{A}_j se označují jako gradientní.

Efektivní jsou metody hybridní využívající jak rychlé konvergence Gauss-Newtonovy metody v oblasti minima, tak zajištěné konvergence gradientních metod v oblastech dále od minima. Nejznámějším představitelům je Marquardtova metoda, které odpovídá přírůstkový vektor

$$\underline{M}_j(\lambda) = (J^T J + \lambda E)^{-1} J^T \underline{e} \quad (22)$$

Je zřejmé, že pro $\lambda = 0$ odpovídá $\underline{M}_j(0)$ směru linearizace \underline{L}_j . Pro $\lambda \rightarrow \infty$ je $\underline{M}_j(\infty)$ ve směru záporného gradientu \underline{A}_j , ale $\|\underline{M}_j(\infty)\| = 0$. S růstem λ se tedy \underline{M}_j vzdaluje od směru \underline{L}_j a blíží se ke směru \underline{A}_j při současném zkrácování délky přírůstkového vektoru. Závislost $\underline{M}_j(\lambda)$ na λ je prostorová křivka (aproximuje se často úsekem spirály).

Program MINOPT využívá strategie "double dog leg", kdy se hledá optimální směrový vektor na straně TG trojúhelníka definovaného vrcholy

$$\underline{O} = \underline{a}^{(j)} \quad \underline{T} = \underline{a}^{(j)} + \alpha_{1j}^* \underline{L}_j \quad \underline{G} = \underline{a}^{(j)} + \alpha_{2j}^* \underline{A}_j$$

Pro směrový vektor tedy platí

$$\underline{D}_j = \mu \underline{T} + (1-\mu) \underline{G} \quad (23)$$

kde $0 \leq \mu \leq 1$ je parametr, jehož hodnota se hledá tak, aby byl přírůstkový vektor vhodný (viz rov. (6)).

Jednotlivé programy využívající těchto strategií se odlišují v realizaci některých dílčích úloh. Je proto běžné, že programy téhož názvu se v praktických aplikacích výrazně liší. Mezi rozhodující z numerického hlediska patří :

- způsob invertace matice H resp. $J^T J$
- řízení velikosti přírůstkového vektoru \underline{D}_j

Je výhodné použít pro invertaci matice H metod využívajících rozkladu na

vlastní čísla a vlastní vektory (resp. přímo SVD matice J). Velikost přírůstkového vektoru se obvykle řídí s ohledem na oblast, ve které je kvadratická aproximace $S(\hat{a})$ ještě přijatelná. Obě tyto strategie využívá program MINOPT.

4. Statistická stránka nelineární regrese

Za předpokladu platnosti normality odhadů (rov. (3)) je poměrně jednoduché určit kovarianční matici odhadů

$$D(\hat{a}) = \hat{\sigma}^2 (J^T J)^{-1} \quad (24)$$

a reziduální rozptyl

$$\hat{\sigma}^2 = S(\hat{a}) / (n - m) \quad (25)$$

Na základě těchto veličin lze snadno určit konfidenční intervaly pro jednotlivé parametry konfidenční oblasti a provádět standardní statistické testy. Velká většina programových paketů obsahuje pouze intervaly spolehlivosti parametrů a kovarianční resp. korelační matici odhadů určenou z rov. (24). Jde vlastně o asymptotické odhady, protože rov. (3) platí jen pro $n \rightarrow \infty$. Pro silně nelineární problémy reálné velikosti vede tento postup k výrazně zkresleným výsledkům. Je proto účelné zkoumat stupeň nelinearity regresního modelu a na jeho základě rozhodnout o použitelnosti rov. (24) a souvisejících postupů statistické analýzy /2/. V programu MINOPT je volen alternativní postup vycházející z toho, že vlivem nelinearity modelu a konečného počtu bodů dochází k vychýlení odhadů \hat{a} . Pro určení velikosti vychýlení $b \approx E(\hat{a} - a)$ navrhli Cook a kol. /6/ postup vycházející z porovnání lineárního a kvadratického rozvoje regresního modelu. Vychýlení b se počítá jednoduše ze vztahu

$$b = (J^T J)^{-1} J^T d \quad (26)$$

kde d je vektor se složkami

$$d_i = -\hat{\sigma}^2 \text{tr} [(J^T J)^{-1} G_i] / 2 \quad (27)$$

Zde $\text{tr}(\cdot)$ označuje stopu matice a G_i je matice druhých derivací modelové funkce v bodě x_i s prvky

$$G_{jk} = \frac{\partial^2 f(x_i, a)}{\partial a_j \partial a_k} \quad j, k = 1, \dots, m \quad (28)$$

Pro praktické účely se počítají relativní vychýlení odhadů $b_{Rj} = (b_j / \hat{a}_j) 10^2$. Orientačně platí, že pro $b_{Rj} < 1$ lze použít klasickou analýzu nelineární regrese využívající asymptotických vztahů. (Je ekvivalentní linearizaci regresního modelu a statistické analýze vzniklého lineárního modelu).

Pokud je vychýlení vyšší, je třeba použít alternativních postupů. Program MINOPT využívá v těchto případech techniku Jackknife ve variantě navržené v práci /7/. Další podrobnosti o statistické analýze použité v programu MINOPT lze nalézt v práci /8/.

5. Charakteristika porovnávaných paketů

Pro porovnání byly vybrány dostupné nejznámější profesionální programové pakety BMDP'87, SAS 6.03, STATGRAPHICS 2.6, SYSTAT 3.0, ASYSTANT+, a SPSS PC + 3.1. Tyto pakety byly doplněny o program MINOPT z našeho balíčku CHEMSTAT. Zde se pouze stručně zmíníme o jednotlivých balíčcích s ohledem na způsob testace.

A. BMDP

Programový balíček BMDP se skládá z izolovaných programů pro různé statistické výpočty. Ve verzi PC jde o 29 programů rozčleněných do 13 bloků.

Programy pro nelineární regresi jsou v bloku regrese.

Vyžaduje operační paměť 512 a u některých programů 640 kbyte, pevný disk a zejména matematický koprocesor. V jednotlivých verzích (86, 87, 88) se liší zejména co do snadnosti zadávání řídicích příkazů. Ve starších verzích se řídicí příkazy sestavovaly do souboru mimo vlastní BMDP. V novějších je řádkový, resp. celoobrazovkový editor.

Vzhledem k tomu, že je původní BMDP vytvořen v jazyce FORTRAN, obsahují řídicí programy poměrně pevnou strukturu s řadou povinných částí (včetně ukončení tečkou). Pro nelineární regresi jsou k dispozici programy 3R a AR. Program 3R umožňuje hledání minima $S(\underline{a})$ pomocí Gauss-Newtonovy metody. Lze použít buď zadaných, nebo zvolených nelineárních funkcí a pracovat s lineárními omezeními na parametry ve tvaru rovnosti.

Program AR využívá nederivační metody DVD (místo tečné roviny definované sloupci matice J se konstruuje sečná rovina). Může pracovat s lineárními omezeními ve tvaru nerovností. Ve verzi BMDP'88 lze použít také modelů zadaných diferenciálními rovnicemi. Ve verzích PC jsou omezeny možnosti využití jiných kritérií regrese.

K testaci byly použity programy 3R a AR z verze BMDP'87.

B. SAS

Rozsáhlý programový soubor SAS je široce využíván zejména u sálových počítačů a minipočítačů. Ve verzi PC využívá ke své práci řady možností od systému nabídek (vyplňování panelu) až k dávkové práci přes řídicí soubory. I když se skládá z řady relativně samostatných modulů, potřebuje ke své činnosti pevný disk. Nevyžaduje však matematický koprocesor. Pro nelineární regresi slouží procedura NLIN. Zde lze odhadovat parametry nelineárních regresních modelů MNČ nebo VMNČ s využitím Gauss-Newtonovy, Marquardtovy, gradientní a nederivační DVD metody. Lze volit i řadu heuristických strategií pro určování optimálních přírůstků (půlení intervalu, zlatý řez, Armijeva metoda, kubická aproximace) a způsob invertace matice $J^T J$ (Moore-Penrosova inverze, pseudo inverze). Opět mají jednotlivé příkazy pevnou strukturu a jsou v povinném tvaru. Je nutné zadávat derivace regresní funkce dle parametrů. Pro testaci byla použita verze 6.03.

C. STATGRAPHICS

Jde o typického představitele statistického balíku konstruovaného speciálně pro PC. Je ovládán hierarchickým systémem nabídek. Alternativně lze však použít také příkazy. Lze ho spouštět i z disket a nevyžaduje matematický koprocesor. Pro nelineární regresi se používá Marquardtova metoda, kterou lze spustit buď ze vstupního panelu, nebo příkazem NONLIN. Výhodou je, že vstupní panel obsahuje předdefinované hodnoty, takže uživatel mění jen to, co potřebuje (nebo zná). Základní nevýhodou je způsob zadávání regresního modelu, který se vyčísľuje zprava doleva bez uvažování obvyklých priorit. Pro testy byla použita verze 2.6.

D. SYSTAT

Tento programový modulární balíček je řadou statistiků považován za jeden z nejvhodnějších. Lze ho provozovat i z disket a nevyžaduje matematický koprocesor. Může pracovat s vlastním jazykem podobným BASICu jak v interaktivním příkazovém, tak i dávkovém režimu.

Pro nelineární regresi je určen modul NONLIN, který umožňuje odhadovat parametry nelineárních regresních modelů při volbě celé řady různých kritérií regrese (včetně MNČ a L_p). Je možné volit derivační metodu variabilní metriky

(Fletcherova varianta) a nederivační simplexovou metodu. Kombinací příkazů MODEL a LOSS lze řešit řadu praktických úloh, jako je vynechávání skupin bodů, úseková regrese, omezení na parametry atd. Pro testaci byla použita verze 3.0.

E. ASYSTANT+

Jde o integrovaný paket zaměřený na technické aplikace. Umožňuje sběr dat, jejich předzpracování a statistickou analýzu. Základem systému je kalkulátor a nabídka členěná hierarchicky. Kalkulátor slouží kromě řady jiných věcí ke specifikaci a zadávání proměnných resp. parametrů. Vyžaduje ke své práci pevný disk a matematický koprocessor. V původní verzi je hardwareově chráněn. V nabídce regresní analýza je blok "Curve fit" pro nelineární regresi. Uživatel vyplňuje panel obsahující zadaný model, první odhady a parametry minimalizace. Je použito pouze MNČ. Zvláštností je, že lze volit, které parametry budou pevné a které se budou měnit. Kvalita prvních odhadů je znázorněna graficky porovnáním průběhu $f(x, a^{(0)})$ s experimentálními body před vlastní minimalizací. Pro minimalizaci lze použít metody Gauss-Newtonovu, variabilní metriky a hybridní kombinace těchto dvou. Pro testaci byla použita verze 1.0.

F. SPSS PC

Jde opět o jednoho z nejznámějších statistických paketů převedených ze sálových počítačů. V minimální verzi vyžaduje kolem 4 Mbyte, ale nevyžaduje matematický koprocessor. Je řešen modulárně a umožňuje práci jak v příkazovém interaktivním módu, tak i v dávkovém módu.

V bloku "Advanced Statistics" obsahuje program pro nelineární regresi, který je opět řízen systémem poměrně pevných příkazů. Lze použít MNČ a Marquardtovu metodu. Pro testaci byla použita verze 3.0.

G. CHEMSTAT

Jde o původní rozsáhlý statistický paket dodávaný družstvem Trilobyte (ve spolupráci s JZD Slušovice). Je složen z několika modulů. Lze ho spouštět také z diskety a nevyžaduje matematický koprocessor. Je plně ovládán hierarchickým systémem nabídek. Uživatel pouze vyplňuje panel se svým zadáním. Lze použít MNČ a VMNČ. Minimalizace se provádí pomocí hybridní strategie (viz kap. 3). Na rozdíl od ostatních paketů obsahuje rozsáhlejší statistickou analýzu (viz kap. 4).

6. Testační studie

Pro porovnání úrovně numerické stránky programů pro nelineární regresi z výše uvedených paketů byla realizována jednoduchá série testů. Testační příklady (v základní verzi 4 a v rozšířené 6) jsou uvedeny v příloze. Z důvodu porovnávání byly všude, kromě programů ze systému SAS 6.03, počítány derivace numericky a byl volen pokud možno dávkový režim.

Pro vyjádření kvality jednotlivých paketů z hlediska schopnosti řešit úlohy nelineární regrese byl určen tzv. index kvality

$$PI = 10^2 \frac{\text{počet správných výsledků}}{T * \text{počet použitých metod}}$$

Zde T je počet použitých testovacích příkladů. V prvním kole byly použity 4 základní testy (T = 4) a ve druhém kole (pouze pro metody úspěšné v prvním kole) další dva testy (T = 6). Výsledky jsou sumarizovány v tabulce 1.

Tabulka 1 : Indexy kvality programů

Paket	P I (T = 4)	P I (T = 6)
BMDP	25	-
SAS	25	-
SYSTAT	37,5	-
STATGRAPHICS	50	-
ASYSTANT	8,3	-
SPSS	100	66,6
CHEMSTAT	100	100

Je patrné, že jasně nejlepší výsledky poskytuje CHEMSTAT obsahující program MINOPT. Na druhém místě je paket SPSS využívající Marquardtovy metody. I když je v některých ostatních paketech obsaženo více metod, ukázalo se, že ani žádná izolovaná metoda neposkytuje vždy korektní výsledky. Naopak došlo k tomu, že různé problémy vyřešily různé metody.

7. Závěr

Na základě výše uvedeného lze konstatovat, že k řešení úloh nelineární regrese pomocí programových paketů na PC je nutné přistupovat obezřetně a nespolehat na to, že počítač je kvalitní a software je od renomované firmy. Zájemcům o CHEMSTAT může podrobnosti poskytnout autor příspěvku.

8. Literatura

- /1/ Militký J. : Tvorba matematických modelů VI, skripta DT Ostrava 1989
- /2/ Militký J. : Tvorba matematických modelů V, skripta DT Ostrava 1988
- /3/ Meloun M., Militký J. Chemometrie IV, skripta pro PGS, VŠCHT Pardubice
- /4/ Demidenko E.Z. : Optimizacija i regressija, Nauka, Moskva 1989
- /6/ Cook R.D. a kol. : Biometrika 73, 615 (1986)
- /7/ Simonoff J.S., Tsai Ch. L. : Technometrics, 28, 103 (1986)
- /8/ Militký J., Čáp J. : Proc. Conf. CEF'87, Taormina Sicilia 1987

Příloha - testovní příklady

Příklad 1 :

Zdroj : Meyer R.R., Roth P.M.: J.I.M.A. 9, 218 (1972)

$$y = a_1 + a_2 \exp(a_3 x)$$

Data :

x	1	5	10	15	20	25	30	35	40	50
y	16.7	16.8	16.9	17.1	17.2	17.4	17.6	17.9	18.1	18.7

První odhady : $a_1^0 = 20$, $a_2^0 = 2$, $a_3^0 = 0.5$

$$S(\underline{a}^0) = 2.1 \cdot 10^{22}$$

Konečné odhady : $\hat{a}_1 = 15.67$, $\hat{a}_2 = 0.999$, $\hat{a}_3 = 0.0222$

$$S(\underline{\hat{a}}) = 5.986 \cdot 10^{-3}$$

Příklad 2 :

Zdroj : Jennrich D.I., Sampson P.F.: Technometrics, 10, 63 (1968)

$$y = \exp(a_1 x) + \exp(a_2 x)$$

Data :

x	1	2	3	4	5	6	7	8	9	10
y	4	6	8	10	12	14	16	18	20	22

První odhady : $a_1^0 = 0.3$ $a_2^0 = 0.4$. $S(\underline{a}^0) = 4.17 \cdot 10^3$

Konečné odhady: $\hat{a}_1 = \hat{a}_2 = 0.2578$ $S(\underline{\hat{a}}) = 124.34$

Příklad 3 :

Zdroj : Meyer R.R., Roth P.M.: J.I.M.A. 9 218 (1972)

$$y = a_1 \exp(a_2 / (a_3 + x))$$

Data :

x	y
50	34 780
55	28 610
60	23 650
65	19 630
70	16 370
75	13 720
80	11 540
85	9 744
90	8 261
95	7 030
100	6 005
105	5 147
110	4 427
115	3 820
120	3 307
125	2 872

První odhady : $a_1^0 = 0.02$ $a_2^0 = 4 000$ $a_3^0 = 250$ $S(\underline{a}^0) = 1.7 \cdot 10^9$

Konečné odhady : $\hat{a}_1 = 0.00562$ $\hat{a}_2 = 6 180$ $\hat{a}_3 = 345.2$ $S(\underline{\hat{a}}) = 87.9$

Příklad 4 :

Zdroj : Květoň K.: Simulovaná data

$$y = a_1 \exp(-a_3 x) + a_2 \exp(-a_4 x)$$

Data :

x	1	2	3	4	5	6	7	8	9	10
y	99,6	67,1	45,9	31,9	22,5	16,1	11,7	8,6	6,38	4,78

První odhady : $a_1^0 = a_2^0 = a_3^0 = a_4^0 = 1$ $S(\underline{a}^0) = 1,84 \cdot 10^4$

Konečné odhady : $\hat{a}_1 = 47,97$ $\hat{a}_2 = 102,051$
 $\hat{a}_3 = 0,2466$ $\hat{a}_4 = 0,4965$ $S(\hat{a}) = 3,179 \cdot 10^{-4}$

Příklad 5 :

Zdroj : Hřebíček J.: Experimentální data

$$y = a_1 \exp(a_3 x) + a_2 \exp(a_4 x)$$

Data :

x	y
7,448	57,544
7,448	53,546
9,439	3,006
7,969	19,498
8,176	16,444
9,284	4,305
7,552	45,290
7,877	27,952
8,552	11,803
9,314	4,764
7,607	51,286
7,847	31,623
8,176	21,777
8,523	13,996
8,903	7,727

První odhady : $a_1^0 = 10^5$ $a_2^0 = 10^5$ $a_3^0 = -1,679$
 $a_4^0 = -1,31$ $S(\underline{a}^0) = 1,12 \cdot 10^4$

Konečné odhady : $\hat{a}_1 = 3,402 \cdot 10^7$ $\hat{a}_2 = 1,651 \cdot 10^3$
 $\hat{a}_3 = -1,816$ $\hat{a}_4 = -0,674$ $S(\hat{a}) = 129$

Příklad 6 :

Zdroj : Květoň K.: Simulovaná data

$$y = a_1 x^{a_3} + a_2 x^{a_4}$$

Data :

x	12	13	14	15	16	17	18	19	20	21	22	23
y	7,31	7,55	7,80	8,05	8,31	8,57	8,84	9,12	9,40	9,69	9,99	10,3

První odhady : $a_1^0 = 100$ $a_2^0 = 0,01$ $a_3^0 = 2$ $a_4^0 = 100$
 $S(\underline{a}^0) = 2,86 \cdot 10^{23}$

Konečné odhady : $\hat{a}_1 = 3,802$ $\hat{a}_2 = 4,14 \cdot 10^{-3}$
 $\hat{a}_3 = 0,223$ $\hat{a}_4 = 2,061$ $S(\hat{a}) = 2,98 \cdot 10^{-5}$