14 Parametrizace v nelineární regresi

V nelineární regresi se setkáme s novým jevem v porovnání s regresí lineární. Když při vyšetřování vlastností odhadů použijeme aproximaci jemnější než lineární, zjistíme, že odhad vektoru parametrů $\boldsymbol{\theta}$ obecně není nestranný a že jeho vychýlení může záviset na tom, jak jsme regresní funkci vyjádřili pomocí parametrů.

14.1 Označení

Rozšiřme označení zavedené v odstavci 13.3. Symbolem $\ddot{\mathbf{F}}(\boldsymbol{\theta})$ označíme trojrozměrné pole typu $n \times k \times k$ dané vztahem

$$\ddot{\mathbf{F}}(\boldsymbol{\theta}) = \frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\mathbf{f}(\boldsymbol{\theta}) \tag{14.1}$$

$$= \left(\ddot{\mathbf{F}}_{i\bullet\bullet}(\boldsymbol{\theta})\right)_{i=1,\dots,n} = \left(\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}f(x_i,\boldsymbol{\theta})\right)_{i=1,\dots,n} \tag{14.2}$$

$$= \left(\ddot{\mathbf{f}}_{\bullet jr}(\boldsymbol{\theta})\right)_{j,r=1,\dots,k} = \left(\frac{\partial^2}{\partial\theta_j\partial\theta_r}\mathbf{f}(\boldsymbol{\theta})\right)_{j,r=1,\dots,k}. \tag{14.3}$$

Pole $\ddot{\mathbf{F}}(\boldsymbol{\theta})$ si můžeme představit jako trojrozměrný objekt se čtvercovou základnou a výškou n, jehož i-tá vrstva je tvořena maticí $\ddot{\mathbf{F}}_{i\bullet\bullet}$ a jr-tý sloupec vektorem $\ddot{\mathbf{f}}_{\bullet jr}$. Podobně jako dříve označíme pole $\ddot{\mathbf{F}}(\boldsymbol{\theta}^*)$ symbolem $\ddot{\mathbf{F}}^*$.

Nyní použijeme kvadratickou aproximaci regresní funkce

$$f(x_i;\boldsymbol{\theta}) \doteq f(x_i;\boldsymbol{\theta}^*) + \frac{\partial}{\partial\boldsymbol{\theta}'}f(x_i;\boldsymbol{\theta}^*)(\boldsymbol{\theta}-\boldsymbol{\theta}^*) + \frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}^*)'\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}f(x_i;\boldsymbol{\theta}^*)(\boldsymbol{\theta}-\boldsymbol{\theta}^*).$$

Uvedenou aproximaci vektoru $\mathbf{f}(\boldsymbol{\theta})$ pro všech n složek vektoru $\mathbf{f}(\boldsymbol{\theta})$ pomocí trojrozměrného pole $\ddot{\mathbf{F}}(\boldsymbol{\theta})$ a s využitím zkráceného zápisu pro hodnoty $\boldsymbol{\theta} = \boldsymbol{\theta}^*$ zapíšeme jako

$$\mathbf{f}(\boldsymbol{\theta}) \doteq \mathbf{f}^* + \mathbf{F}^*(\boldsymbol{\theta}-\boldsymbol{\theta}^*) + \frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}^*)'\,\ddot{\mathbf{F}}^*(\boldsymbol{\theta}-\boldsymbol{\theta}^*)\,. \tag{14.4}$$

Jde o rozšíření lineární aproximace (13.10) o kvadratický člen, v němž se při násobení trojrozměrné matice $\ddot{\mathbf{F}}(\boldsymbol{\theta})$ provádí součin přes druhý a třetí rozměr.

Ukažme si budoucí problémy na dvou velmi jednoduchých příkladech.

Příklad 14.1 Mějme regresní funkci $f(x,\theta) = \mathrm{e}^{x\theta}$. Zavedeme-li nový parametr $\beta = \mathrm{e}^{\theta}$, můžeme stejnou funkci proměnné x zapsat jako $\tilde{f}(x,\beta) = \beta^x$. Je tedy

$$\begin{aligned}
\frac{\partial}{\partial\theta} f(x,\theta) = x\mathrm{e}^{x\theta} &= xf(x,\theta),\\
\frac{\partial^2}{\partial\theta^2} f(x,\theta) = x^2\mathrm{e}^{x\theta} &= x^2 f(x,\theta),\\
\frac{\partial}{\partial\beta} \tilde{f}(x,\beta) = x\beta^{x-1} &= \frac{x}{\beta}\tilde{f}(x,\beta),\\
\frac{\partial^2}{\partial\beta^2} \tilde{f}(x,\beta) = x(x-1)\beta^{x-2} &= \frac{x(x-1)}{\beta^2}\tilde{f}(x,\beta).
\end{aligned}$$

Pozorování provedeme pro $x_1 = 0$, $x_2 = 1$, takže bude

$$\begin{aligned}
\mathbf{f}(\theta) &= \begin{pmatrix} 1 \\ \mathrm{e}^{\theta} \end{pmatrix}, & \mathbf{F}(\theta) &= \begin{pmatrix} 0 \\ \mathrm{e}^{\theta} \end{pmatrix}, & \ddot{\mathbf{F}}(\theta) &= \begin{pmatrix} 0 \\ \mathrm{e}^{\theta} \end{pmatrix},\\
\tilde{\mathbf{f}}(\beta) &= \begin{pmatrix} 1 \\ \beta \end{pmatrix}, & \tilde{\mathbf{F}}(\beta) &= \begin{pmatrix} 0 \\ 1 \end{pmatrix}, & \ddot{\tilde{\mathbf{F}}}(\beta) &= \begin{pmatrix} 0 \\ 0 \end{pmatrix}.
\end{aligned}$$

Zvolíme $\theta^\star = 0$, čemuž odpovídá $\beta^\star = 1$. Výsledné *lineární* aproximace jsou

$$\begin{aligned}
\mathbf{f}(\theta) &= \begin{pmatrix} 1 \\ \mathrm{e}^{\theta} \end{pmatrix} \doteq \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix}\theta = \begin{pmatrix} 1 \\ 1+\theta \end{pmatrix},\\
\tilde{\mathbf{f}}(\beta) &= \begin{pmatrix} 1 \\ \beta \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix}(\beta-1) = \begin{pmatrix} 1 \\ \beta \end{pmatrix}.
\end{aligned}$$

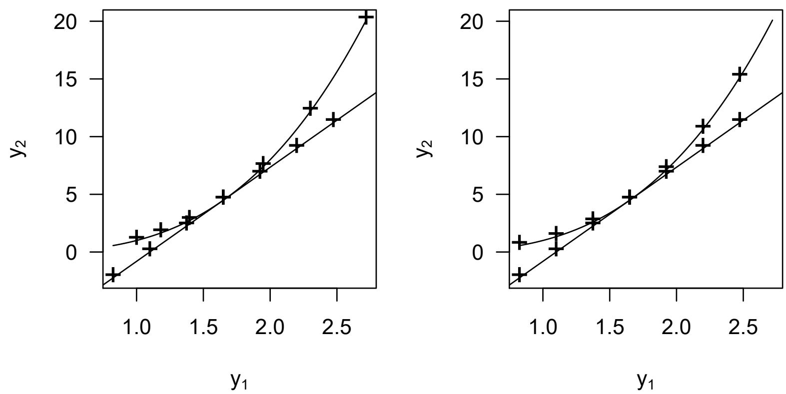
Zatímco v prvním případě jde skutečně o aproximaci, ve druhém případě máme místo aproximace identitu. Zvolíme-li $\theta \neq 0$, budou vektory $\mathbf{f}(\theta)$ a jeho aproximace neshodné. Při rovnoměrném měnění θ probíhá příslušný bod v aproximaci rovnoměrně, kdežto skutečný bod $\mathbf{f}(\theta)$ mění svoji rychlost. ○

Příklad 14.2 Zvolme nyní pro stejné funkce f, $\tilde{f}$ jako v příkladu 14.1, ale $x_1 = 1$ a $x_2 = 3$, dostaneme poněkud jiné matice

$$\begin{aligned}
\mathbf{f}(\theta) &= \begin{pmatrix} e^{\theta} \\ \mathrm{e}^{3\theta} \end{pmatrix}, & \mathbf{F}(\theta) &= \begin{pmatrix} e^{\theta} \\ 3\mathrm{e}^{3\theta} \end{pmatrix}, & \ddot{\mathbf{F}}(\theta) &= \begin{pmatrix} e^{\theta} \\ 9\mathrm{e}^{3\theta} \end{pmatrix},\\
\tilde{\mathbf{f}}(\beta) &= \begin{pmatrix} \beta \\ \beta^3 \end{pmatrix}, & \tilde{\mathbf{F}}(\beta) &= \begin{pmatrix} 1 \\ 3\beta^2 \end{pmatrix}, & \ddot{\tilde{\mathbf{F}}}(\beta) &= \begin{pmatrix} 0 \\ 6\beta \end{pmatrix}
\end{aligned}$$

a také jiné aproximace

$$\begin{aligned}
\mathbf{f}(\theta) &= \begin{pmatrix} \mathrm{e}^{\theta} \\ \mathrm{e}^{3\theta} \end{pmatrix} \doteq \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ 3 \end{pmatrix}\theta = \begin{pmatrix} 1+\theta \\ 1+3\theta \end{pmatrix},\\
\tilde{\mathbf{f}}(\beta) &= \begin{pmatrix} \beta \\ \beta^3 \end{pmatrix} \doteq \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ 3 \end{pmatrix}(\beta-1) = \begin{pmatrix} \beta \\ 3\beta-2 \end{pmatrix}.
\end{aligned}$$



Obrázek 14.1: Množiny možných středních hodnot z příkladu 14.2 a jejich tečné přímky (parametry: vlevo θ, vpravo β)

Na obrázku 14.1 jsou znázorněny části množin možných středních hodnot. Zvlášť jsou vyznačeny střední hodnoty pro lineárně se měnící hodnoty parametru θ (vlevo sedm hodnot od 0 do 1) a β (vpravo sedm hodnot od $\mathrm{e}^0 = 1$ do $\mathrm{e}^1 = \mathrm{e}$). Velikost kroku této změny byla zvolena tak, aby lineární aproximace (body na tečné nadrovině s dotykem v bodě $\mathbf{f}(0{,}5) = \tilde{\mathbf{f}}(\sqrt{\mathrm{e}})$) byly pro obě parametrizace identické. Na druhé straně odstupy odpovídajících bodů na grafu funkce nejsou stejné, v parametrizaci pomocí β jsou rovnoměrnější. Dále je zajímavé porovnat, jak si navzájem odpovídají dvojice bodů na křivce (množina možných středních hodnot) a na aproximující přímce. Parametrizace pomocí β vypadá lépe, vzdálenosti mezi sobě odpovídajícími body nejsou tak veliké. ○

14.2 Odhad vychýlení

Nyní se pokusíme vyjádřit vychýlení odhadu $\mathbf{t}$. Učiníme to nepřímo tak, že porovnáme lineární a kvadratickou aproximaci.

Další postup založíme na následujícím zjednodušujícím předpokladu, který je v použité aproximaci splněn: *Střední hodnota průmětu vektoru* $\mathbf{f}(\mathbf{t})$ *do tečné nadroviny k množině možných středních hodnot v bodě* $\mathbf{f}(\boldsymbol{\theta}^*)$ *je rovna* $\mathbf{f}(\boldsymbol{\theta}^*)$. Podrobné odvození lze nalézt v Box (1971).