Нелинейный метод наименьших квадратов

Дмитрий Алексеевич Павлов
инж.-иссл. Лаборатории Чебышёва СПбГУ
d.a.pavlov@spbu.ru

СПбГУ, МКН
2021

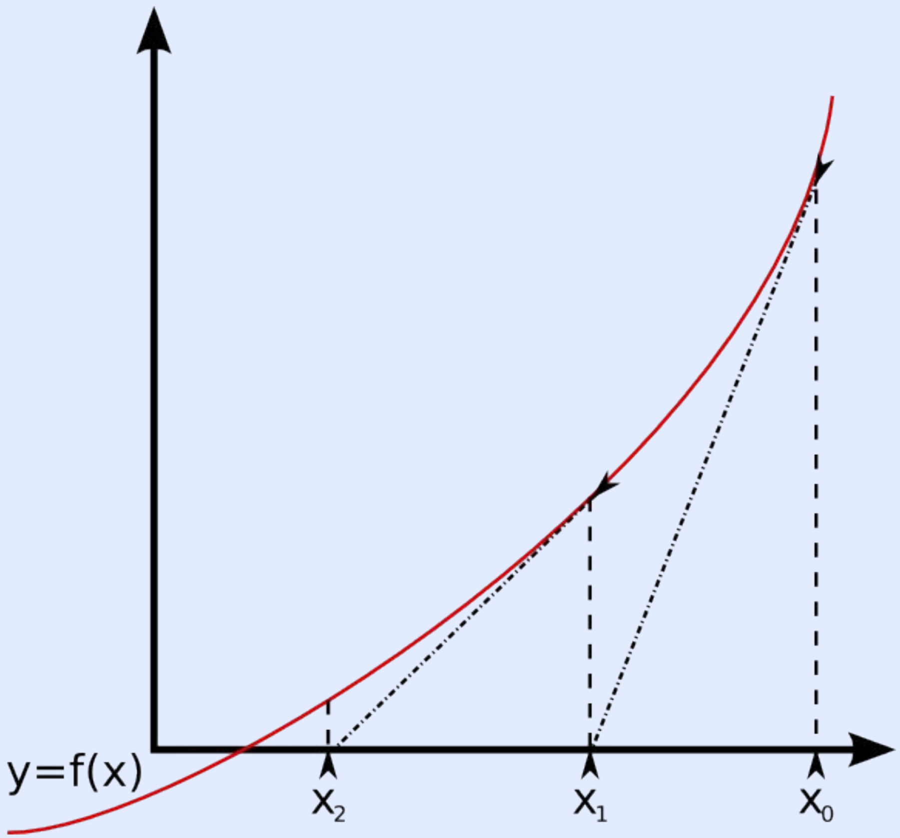
Метод Ньютона

Найти $x^* \ : \ f(x^*) = 0$

$x^{(0)}$ – *начальное приближение.*

$$x^{(l+1)} = x^{(l)} - \frac{f(x^{(l)})}{f'(x^{(l)})}$$

Функция f *линеаризуется* в окрестности $x^{(l)}$.



Метод Ньютона для функции многих переменных

Найти $\mathbf{x}^* \; : \; \mathbf{f}(\mathbf{x}^*) = \mathbf{0}$

$\boldsymbol{x}^{(0)}$ – *начальное приближение.*

$$\mathbf{x}^{(l+1)} = \mathbf{x}^{(l)} - \nabla^{-1}\mathbf{f}(\mathbf{x}^{(l)})$$

Функция **f** всё так же *линеаризуется* в окрестности $\boldsymbol{x}^{(l)}$.

$$\nabla\mathbf{f} = \left[\frac{\partial\mathbf{f}}{\partial x_1} \quad \cdots \quad \frac{\partial\mathbf{f}}{\partial x_n}\right] = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix}$$ – якобиан

Нелинейная регрессия

Модель

$$y = g(x_1, \ldots, x_k, \beta_1, \ldots \beta_m) + \varepsilon$$

y – зависимая переменная (наблюдаемая)

x_i – влияющие переменные (наблюдаемые)

β_j – параметры модели (ненаблюдаемые)

ε – случайные ошибки (шум)

Наблюдения

$$y_i = g(x_{i1}, \ldots, x_{ik}, \beta_1, \ldots \beta_m) + \varepsilon_i$$

$$\mathrm{Var}[\varepsilon_i] = \sigma_i^2$$

Нелинейный взвешенный МНК

Дано

$\mathbf{y} = (y_1, y_2, \ldots, y_n)^T$ – наблюдения

$$\begin{pmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nk} \end{pmatrix}$$

Найти $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_m)$:

$$y_i = g(x_{i1}, \ldots, x_{ik}, \hat{\beta}_1, \ldots \hat{\beta}_m) + r_i$$

$$\arg\min_{\hat{\boldsymbol{\beta}}} S(\hat{\boldsymbol{\beta}}) = ?$$

$$S(\hat{\boldsymbol{\beta}}) = \sum_{j=1}^{m} \frac{r_j^2(\hat{\boldsymbol{\beta}})}{\sigma_i^2}$$

Оценка не является несмещённой.

Состоятельность доказана при выполнении некоторых условий.

(Но никого это не интересует.)

Алгоритм Гаусса-Ньютона

$$\nabla S(\hat{\boldsymbol{\beta}}) = \mathbf{0} \Rightarrow f_j(\hat{\boldsymbol{\beta}}) = \sum_{i=1}^{n} \frac{\partial r_i(\hat{\boldsymbol{\beta}})}{\partial \hat{\beta}_j} \frac{r_i(\hat{\boldsymbol{\beta}})}{\sigma_i^2} = 0, \quad j = 1..m$$

$$A = \begin{bmatrix} \frac{\partial r_1}{\partial \hat{\beta}_1} & \cdots & \frac{\partial r_1}{\partial \hat{\beta}_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial r_n}{\partial \hat{\beta}_1} & \cdots & \frac{\partial r_n}{\partial \hat{\beta}_m} \end{bmatrix} \quad W = \begin{bmatrix} \frac{1}{\sigma_1^2} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\sigma_n^2} \end{bmatrix}$$

$$\mathbf{f}(\hat{\boldsymbol{\beta}}) = A^T W \mathbf{r}(\hat{\boldsymbol{\beta}}) = 0$$

$$\frac{\partial f_j(\hat{\boldsymbol{\beta}})}{\partial \hat{\beta}_l} = \sum_{i=1}^{n} \frac{1}{\sigma_i^2} \left(\frac{\partial r_i(\hat{\boldsymbol{\beta}})}{\partial \hat{\beta}_j} \frac{\partial r_i(\hat{\boldsymbol{\beta}})}{\partial \hat{\beta}_l} + r_i \frac{\partial^2 r_i(\hat{\boldsymbol{\beta}})}{\partial \hat{\beta}_j \partial \hat{\beta}_l} \right) \approx \sum_{i=1}^{n} \frac{1}{\sigma_i^2} \frac{\partial r_i(\hat{\boldsymbol{\beta}})}{\partial \hat{\beta}_j} \frac{\partial r_i(\hat{\boldsymbol{\beta}})}{\partial \hat{\beta}_l}$$

$$\nabla \mathbf{f} \approx A^{\mathrm{T}} W A \qquad \hat{\boldsymbol{\beta}}^{(l+1)} = \hat{\boldsymbol{\beta}}^{(l)} - (A^{\mathrm{T}} W A)^{-1} A^{\mathrm{T}} W \mathbf{r}(\hat{\boldsymbol{\beta}}^{(l)})$$

Определение параметров динамической системы

$$(x_{i1}, \ldots, x_{ik}) \equiv \mathbf{x}_i = \mathbf{x}(t_i)$$

$$y_i = g(\mathbf{x}(t_i); \mathbf{q}) + \varepsilon_i$$

$$\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}(t); \mathbf{p})$$

$$\mathbf{x}(t_0) = \mathbf{x}_0$$

$$\boldsymbol{\beta} = (\mathbf{x}_0 \mid \mathbf{p} \mid \mathbf{q})$$

$\mathbf{x}_0$ – начальные условия
p – динамические параметры
q – редукционные параметры

$$\frac{\partial r_i}{\partial \mathbf{q}} = -\frac{\partial g(\mathbf{x}(t_i); \mathbf{q})}{\partial \mathbf{q}}$$

$$\frac{\partial r_i}{\partial (\mathbf{x}_0 \mid \mathbf{p})} = -\frac{\partial g(\mathbf{x}(t_i); \mathbf{q})}{\partial \mathbf{x}} \frac{\partial \mathbf{x}(t_i)}{\partial (\mathbf{x}_0 \mid \mathbf{p})}$$

Изохронные производные

$$\frac{\partial \mathbf{x}(t)}{\partial(\mathbf{x}_0 \mid \mathbf{p})} = ?$$

(Можно рассчитать численно с инкрементом параметра, но это уныло.)

$$\mathbf{P} \equiv (\mathbf{x}_0 \mid \mathbf{p})$$

$$\left(\frac{\partial \mathbf{x}}{\partial \mathbf{P}}\right)^{\cdot} = \frac{\partial \mathbf{F}}{\partial \mathbf{P}} = \frac{\partial \mathbf{F}}{\partial \mathbf{x}} \frac{\partial \mathbf{x}}{\partial \mathbf{P}}$$

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}\mathbf{P}}(t_0) = \begin{array}{c} \begin{bmatrix} 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 & \cdots & 0 \end{bmatrix} \\ \mathbf{x}_0 \qquad\qquad\qquad \mathbf{p} \end{array}$$

$\partial\mathbf{x}/\partial\mathbf{P}$ делается частью динамической системы.

Что дальше?

- Обратите внимание, что функция g не обязана быть для всех i одной и той же. Можно обрабатывать разные типы наблюдаемых величин в одном нелинейном МНК.
- Для решения задачи Коши, т. е. вычисления $\mathbf{x}(t)$, см. методы численного интегрирования, начиная с метода Рунге-Кутты 4-го порядка.