Programowanie narzędzi analitycznych Z06

Rafał Woźniak

Faculty of Economic Sciences, University of Warsaw

Warszawa, 18-11-2021

Metoda Newtona

Metoda Newtona do znajdowania miejsca zerowego

$$\begin{aligned} y - y_0 &= f'(x)(x - x_0) \\ f'(x_0) &= \frac{f(x_0) - 0}{x - x_0} \\ x_{n+1} &= x_n - \frac{f(x_n)}{f'(x_n)} \end{aligned}$$

W maksymalizacji funkcji metodą Newtona znajduje się miejsce zerowe pochodnej

$$x_{n+1} = x_n - \frac{f'(x_n)}{f''(x_n)}$$

Metoda Newtona - maksymalizacja funkcji 2 zmiennych

Niech $\ln L(\theta)$ będzie funkcją log-wiarygodności od argumentu θ, który jest $(K \times 1)$ wektorem.

$$\theta_{n+1} = \theta_n - H_n^{-1} G_n \tag{1}$$

$$G_n = \left. \frac{\partial \ln L(\theta)}{\partial \theta} \right|_{\theta_n} \qquad H_n = \left. \frac{\partial^2 \ln L(\theta)}{\partial \theta \partial \theta'} \right|_{\theta_n} \tag{2}$$

$$G_n = \begin{bmatrix} \frac{\partial \ln L(\theta)}{\partial \theta_1} \\ \frac{\partial \ln L(\theta)}{\partial \theta_2} \\ \vdots \\ \frac{\partial \ln L(\theta)}{\partial \theta_K} \end{bmatrix} \quad H_n = \begin{bmatrix} \frac{\partial^2 \ln L(\theta)}{\partial \theta_1 \partial \theta_1} & \frac{\partial^2 \ln L(\theta)}{\partial \theta_1 \partial \theta_2} & \cdots & \frac{\partial^2 \ln L(\theta)}{\partial \theta_1 \partial \theta_K} \\ \frac{\partial^2 \ln L(\theta)}{\partial \theta_2 \partial \theta_1} & \frac{\partial^2 \ln L(\theta)}{\partial \theta_2 \partial \theta_2} & \cdots & \frac{\partial^2 \ln L(\theta)}{\partial \theta_2 \partial \theta_K} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^2 \ln L(\theta)}{\partial \theta_K \partial \theta_1} & \frac{\partial^2 \ln L(\theta)}{\partial \theta_K \partial \theta_2} & \cdots & \frac{\partial^2 \ln L(\theta)}{\partial \theta_K \partial \theta_K} \end{bmatrix} \tag{3}$$

Metoda Newtona - Przykład z [2]

$$
\begin{aligned}
\ln L(x;\mu,\sigma) &= -\frac{1}{2}N\ln(2\pi) - N\ln(\sigma) - \frac{1}{2}\sum_{i=1}^{N}\frac{(x_i-\mu)^2}{\sigma^2} \\
\frac{\partial}{\partial\mu}\ln L(x;\mu,\sigma) &= \sum_{i=1}^{N}\frac{(x_i-\mu)}{\sigma^2} \\
\frac{\partial}{\partial\sigma}\ln L(x;\mu,\sigma) &= -\frac{N}{\sigma} + \sum_{i=1}^{N}\frac{(x_i-\mu)^2}{\sigma^3} \\
\frac{\partial^2\ln L(x;\mu,\sigma)}{(\partial\mu)^2} &= -\frac{N}{\sigma^2} \\
\frac{\partial^2\ln L(x;\mu,\sigma)}{\partial\mu\partial\sigma} &= -2\sum_{i=1}^{N}\frac{(x_i-\mu)}{\sigma^3} \\
\frac{\partial^2\ln L(x;\mu,\sigma)}{(\partial\sigma)^2} &= \frac{N}{\sigma^3} - 3\sum_{i=1}^{N}\frac{(x_i-\mu)^2}{\sigma^4}
\end{aligned}
$$

Algorytm BHHH

Berndt, E.; Hall, B.; Hall, R.; Hausman, J. (1974). *Estimation and Inference in Nonlinear Structural Models*, Annals of Economic and Social Measurement, 3: 653–665.

$$\theta_{n+1} = \theta_n - J_n^{-1} G_n \tag{4}$$

$$G_n = \begin{bmatrix} \frac{\partial \ln L_1(\theta_n)}{\partial \theta_1} & \frac{\partial \ln L_1(\theta_n)}{\partial \theta_2} & \cdots & \frac{\partial \ln L_1(\theta_n)}{\partial \theta_K} \\ \frac{\partial \ln L_2(\theta_n)}{\partial \theta_1} & \frac{\partial \ln L_2(\theta_n)}{\partial \theta_2} & \cdots & \frac{\partial \ln L_2(\theta_n)}{\partial \theta_K} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial \ln L_N(\theta_n)}{\partial \theta_1} & \frac{\partial \ln L_N(\theta_n)}{\partial \theta_2} & \cdots & \frac{\partial \ln L_N(\theta_n)}{\partial \theta_K} \end{bmatrix} \qquad J_n = G_n G_n' \tag{5}$$

Algorytm BHHH - przykład

$$
\begin{aligned}
f(x;k,\lambda) &= \frac{k}{\lambda^k} x^{k-1} e^{-(x/\lambda)^k}, \qquad x \geqslant 0 \\
\ln L_i(\theta) &= \ln(k) - k\ln(\lambda) + (k-1)\ln(x_i) - \left(\frac{x_i}{\lambda}\right)^k
\end{aligned}
$$

$$
\theta = \begin{bmatrix} k \\ \lambda \end{bmatrix} \qquad G_n(\theta) = \begin{bmatrix} \frac{\partial \ln L_1(\theta_n)}{\partial k} & \frac{\partial \ln L_1(\theta_n)}{\partial \lambda} \\ \frac{\partial \ln L_2(\theta_n)}{\partial k} & \frac{\partial \ln L_2(\theta_n)}{\partial \lambda} \\ \vdots & \vdots \\ \frac{\partial \ln L_N(\theta_n)}{\partial k} & \frac{\partial \ln L_N(\theta_n)}{\partial \lambda} \end{bmatrix}
$$

$$
J_n = \begin{bmatrix} \sum_{i=1}^N \left(\frac{\partial \ln L_i(\theta_n)}{\partial k}\right)^2 & \sum_{i=1}^N \left(\frac{\partial \ln L_i(\theta_n)}{\partial k}\right)\left(\frac{\partial \ln L_i(\theta_n)}{\partial \lambda}\right) \\ \sum_{i=1}^N \left(\frac{\partial \ln L_i(\theta_n)}{\partial k}\right)\left(\frac{\partial \ln L_i(\theta_n)}{\partial \lambda}\right) & \sum_{i=1}^N \left(\frac{\partial \ln L_i(\theta_n)}{\partial \lambda}\right)^2 \end{bmatrix}
$$

Algorytm BHHH - za [1] str. 98

A useful way to think about the structure of the BHHH algorithm is as follows. Let the $(N \times K)$ matrix, X, and the $(N \times 1)$ vector, Y, be given by

$$X = \begin{bmatrix} \frac{\partial \ln L_1(\theta_n)}{\partial k} & \frac{\partial \ln L_1(\theta_n)}{\partial \lambda} \\ \frac{\partial \ln L_2(\theta_n)}{\partial k} & \frac{\partial \ln L_2(\theta_n)}{\partial \lambda} \\ \vdots & \vdots \\ \frac{\partial \ln L_N(\theta_n)}{\partial k} & \frac{\partial \ln L_N(\theta_n)}{\partial \lambda} \end{bmatrix}, Y = \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}. \tag{6}$$

An iteration of the BHHH algorithm is now written as

$$\theta_{n+1} = \theta_n + \left(X_n' X_n\right)^{-1} X_n' Y, \tag{7}$$

where

$$J_n = X_n' X_n, \qquad G_n = X_n' Y. \tag{8}$$

Literatura

[1] V.L. Martin, A.S. Hurn and D. Harris, *Econometric Modelling with Time Series: Specification, Estimation and Testing*, Cambridge University Press, 2012.

[2] Henningsen, Arne and Toomet, Ott (2011). *maxLik: A package for maximum likelihood estimation in R*. Computational Statistics 26(3), 443-458. DOI 10.1007/s00180-010-0217-1. (Link)