Chapter 4

Nonlinear model

4.1 Introduction

Nonlinear model (NLM, *epälineaarinen malli*) is an extension to linear model where the systematic part of the model is no longer a linear function $\mathbf{X}\boldsymbol{\beta}$. Generally, NLM is of form

$$Y_i = \mathrm{f}(x_{i1}, \ldots, x_{ik}; \beta_1, \ldots, \beta_p) + \epsilon_i = \mathrm{f}(\boldsymbol{x}_i; \boldsymbol{\beta}) + \epsilon_i \tag{4.1}$$

for $i = 1, \ldots, n$ observations, k variables and p parameters. Note that for LM $k = p$, but this is not requirement in NLM. In vector form the NLM is

$$\boldsymbol{Y} = \mathbf{f}(\mathbf{X}; \boldsymbol{\beta}) + \boldsymbol{\epsilon}, \tag{4.2}$$

where $\boldsymbol{Y}$ is $n\times 1$, $\mathbf{X}$ is $n\times k$, $\boldsymbol{\beta}$ $p\times 1$, and $\boldsymbol{\epsilon}$ $n\times 1$. Function $\mathbf{f}$ is vector-valued function $(\mathrm{f}(\boldsymbol{x}_1; \boldsymbol{\beta}), \ldots, \mathrm{f}(\boldsymbol{x}_n; \boldsymbol{\beta}))$. In what follows we might shorten $\mathrm{f}(\boldsymbol{x}_i; \boldsymbol{\beta})$ to $\mathrm{f}_i(\boldsymbol{\beta})$ or even to f_i.

4.1.1 Some nonlinear models

Some nonlinear model types are introduced here, but because any (non)linear function f will introduce NLM, the list is merely just a small set of examples. First of all, multiplicative model is NLM if errors are additive, i.e.

$$Y_i = \beta_0 x_{i1}^{\beta_1} \cdots x_{ik}^{\beta_k} + \epsilon_i \tag{4.3}$$

Please note that if errors are also multiplicative, the model can be transformed into linear:

$$Y_i = \beta_0 x_{i1}^{\beta_1} \cdots x_{ik}^{\beta_k} e^{\epsilon_i} \quad \Rightarrow \tag{4.4}$$

$$\log(Y_i) = \log(\beta_0) + \beta_1 \log(x_{i1}) + \ldots + \beta_k \log(x_{ik}) + \epsilon_i \tag{4.5}$$

In modeling the degree of linear polarization in atmosphereless Solar System targets such as asteroids covered with regolith, or dust in comets coma, the so-called trigonometric model is used. It is defined as

$$Y_i = \beta_1 \, \sin(x_i)^{\beta_2} \, \cos(x_i/2)^{\beta_3} \, \sin(x_i - \beta_4) + \epsilon_i, \tag{4.6}$$

where x_i is the phase angle and Y_i is the degree of linear polarization. The function is shown in Fig. 4.1(a).

A model for limited growth is shown in Fig. 4.1(b). The model is

$$Y_i = \beta_1 + \beta_2 \left(1 - e^{-\beta_3 x_i}\right) + \epsilon_i, \tag{4.7}$$

The growth starts from β_1 and is limited by $\beta_1 + \beta_2$. The parameter β_3 controls the speed of growth.

A growth curve can be defined so that it will reach its maximum, but slowly decline after that. A model that is shown in Fig. 4.1(c) is

$$Y_i = \beta_1 + \frac{\beta_2 x_i}{\beta_3 + x_i + \beta_4 x_i^2} + \epsilon_i, \tag{4.8}$$

The growth starts from β_1 and reaches its maximum at $\sqrt{\beta_3/\beta_4}$, but will then decrease.

One more type of growth curves is the S-type curves such as the logistic function in Fig. 4.1(d):

$$Y_i = \frac{\beta_1}{1 + e^{-\beta_2(x_i - \beta_3)}} + \epsilon_i, \tag{4.9}$$

where β_1 controls the limiting value of the growth, β_2 its steepness, and β_3 the location where positive derivative turns into negative.

Many of the NLM's can be derived as a solution for differential equation, for example the growth curves (b) and (d).

4.2 Model estimation

Most of the model estimation and diagnostics are done more or less the same way as in linear model. The main difference is, that results regarding the distribution of parameters, i.e. parameter errors, are always asymptotic, and that the model estimation is a numerical optimization problem. With LM the model estimate is given in closed form, and results regarding parameter distributions are exact under the normal assumption.

Let us derive the NLM parameter estimate from the maximum likelihood principle, although the same result can be reached from the 'minimal least squares' principle. Our model, now with normal assumption, is that

$$\begin{gathered} \epsilon_i \perp\!\!\!\perp \epsilon_j, \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2) \text{ , or alternatively} \\ Y_i \perp\!\!\!\perp Y_j, \quad Y_i \sim \mathcal{N}(\mathrm{f}_i(\boldsymbol{\beta}), \sigma^2) \end{gathered} \tag{4.10}$$

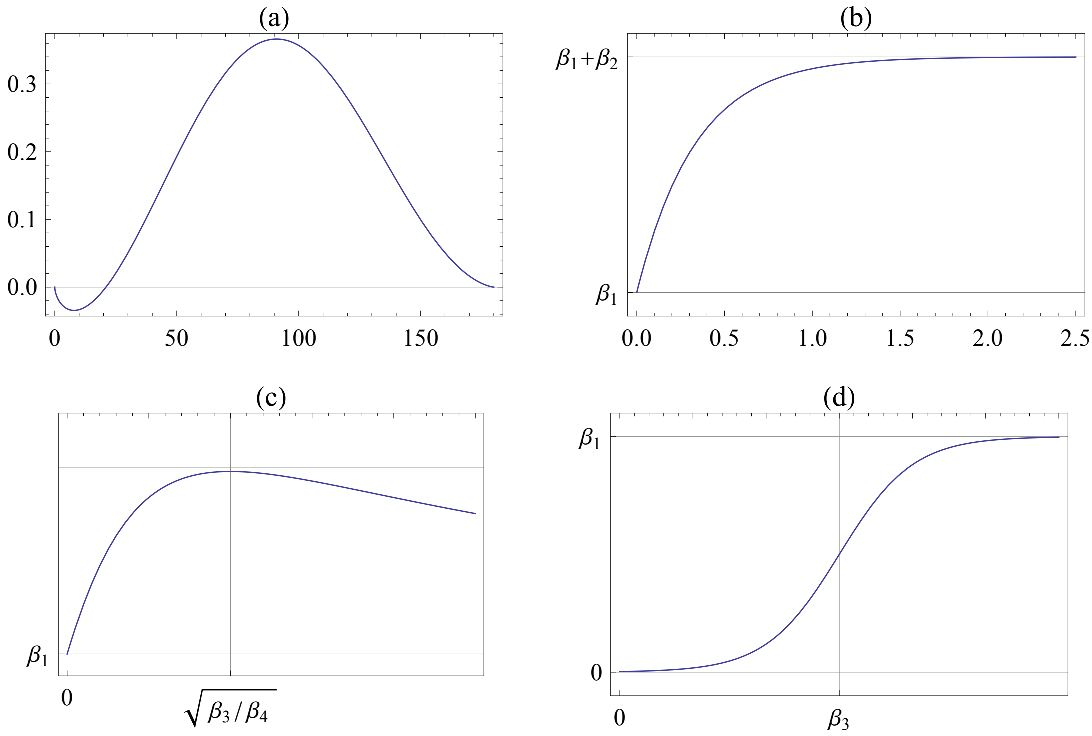

Figure 4.1: Four examples of different models in nonlinear regression.

Because the i.i.d observations, the likelihood function for the model is

$$\mathrm{L}(\boldsymbol{\beta},\sigma^2) = (2\pi\sigma^2)^{-n/2}\,\exp\left(-\frac{1}{2\sigma^2}\sum^{n}\left(y_i - \mathrm{f}_i(\boldsymbol{\beta})\right)^2\right) \tag{4.11}$$

We will write the squared residual sum in a shorter form, $S(\boldsymbol{\beta}) = \sum^n \left(y_i - \mathrm{f}_i(\boldsymbol{\beta})\right)^2$, and state that the log-likelihood function for the model is

$$\mathrm{l}(\boldsymbol{\beta},\sigma^2) = -\frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}S(\boldsymbol{\beta}) \tag{4.12}$$

The maximum of the log-likelihood gives the ML estimates for the NLM. Regarding to parameter vector $\boldsymbol{\beta}$, we can easily see that estimate $\boldsymbol{b} = \hat{\boldsymbol{\beta}}$ must minimize the sum of squared residuals $S(\boldsymbol{\beta})$. When inputting that back to log-likelihood, derivating with respect to σ^2, and searching for root, we find that $s^2 = \hat{\sigma}^2 = \frac{1}{n}S(\boldsymbol{b})$.

Contrary to linear model, the estimate $\boldsymbol{b}$ cannot (usually) be expressed in closed form. The minimization of $S(\boldsymbol{\beta})$ must be done numerically. Quite generally a Gauss-Newton or Levenberg-Marquardt algorithms are used.

4.2.1 Parameter properties

The asymptotic properties of the NLM estimates $\boldsymbol{b}$ and s^2 can be found by analyzing the Hessian matrix of the MLE's (see Eq. (2.6)). After some cumbersome calculus, we can find that for the residual variance we have

$$s^2 \overset{as.}{\sim} \mathcal{N}\left(\sigma^2, \frac{2\sigma^4}{n}\right), \tag{4.13}$$

and for the actual parameters

$$\boldsymbol{b} \overset{as.}{\sim} \mathcal{N}_n\left(\boldsymbol{\beta}, \sigma^2 \left(\mathbf{F}(\boldsymbol{\beta})^T \mathbf{F}(\boldsymbol{\beta})\right)^{-1}\right). \tag{4.14}$$

The matrix $\mathbf{F}(\boldsymbol{\beta})$ is short for the $n \times p$ partial derivative matrix with elements

$$\mathbf{F}(\boldsymbol{\beta}) = \left[\frac{\partial \mathrm{f}_i(\boldsymbol{\beta})}{\partial \beta_j}\right]_{ij}. \tag{4.15}$$

The tests regarding individual parameters in NLM are done in similar manner than with LM, only change being that instead of matrix $\mathbf{M}^{-1} = (\mathbf{X}^T\mathbf{X})^{-1}$ in LM (see Eqs. (3.33)–(3.35)) we have matrix $\mathbf{M}^{-1} = (\mathbf{F}(\boldsymbol{\beta})^T\mathbf{F}(\boldsymbol{\beta}))^{-1}$ in NLM.

The model diagnostics with e.g. residual plots are also done as with LM. The covariance matrix of $\boldsymbol{b}$ is even more important than with LM — highly correlated parameters are hard to estimate with numerical methods. Moving to a different parametrization might help.