8 Nonlinear Regression

Nonlinear regression relates to models, where the mean response is not linear in the parameters of the model.
A MLRM

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

has mean

$$E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k.$$

Even if the regressors x_i are transformations of the explanatory variables like $x_i = z_i^2$, or $x_i = log(z_i)$, the mean response is still linear in $\beta_1, \ldots, \beta_k$.
In nonlinear regression we consider models where the mean response is no longer linear in the model parameters.

We will still assume that the error is added to the mean, so in general the non-linear regression model has the following structure

$$Y = f(\vec{x}, \vec{\beta}) + \varepsilon$$

with $E(\varepsilon) = 0$. Therefore $E(Y) = f(\vec{x}, \vec{\beta})$, and f is the expectation or mean function of the model.
For example

$$y = \theta_1 e^{\theta_2 x} + \varepsilon$$

has mean function $f(x; \theta_1, \theta_2) = \theta_1 e^{\theta_2 x}$, which is not linear in θ_1 and θ_2.

Definition 8.1.
Random variable Y fits a nonlinear regression model (NLRM) if there exist a function $f : \mathbb{R}^p \times \Theta \to \mathbb{R}$ with

$$Y = f(\vec{x}, \vec{\beta}) + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

If f is linear in the parameter vector, the MLRM becomes a special case of the NLRM.

Some nonlinear models origin from the analysis of scatterplots, which might indicate that only a nonlinear model can be fit to the data.
But quite often a nonlinear model arises from the insight into the subject matter for example a certain relationship might arise as a solution of a differential equation.

Example 8.1.
(see MPV pg. 392) Incorporation of temperature into a second order kinetics model. Hydrolysis of ethyl acetate is well modeled by a second order kinetics model. A_t is the amount of ethyl acetate at time t, then

$$\frac{dA_t}{dt} = -kA_t^2$$

with k being the rate constant, which depends on the temperature.
With A_0 the amount of ethyl acetate at time 0, then the solution of the above differential equation is

$$\frac{1}{A_t} = \frac{1}{A_0} + kt$$

which is equivalent to

$$A_t = \frac{A_0}{1 + A_0 kt}$$

The Arrhenius equation relates the reaction rate with the temperature, it is an empirical law which is consistent with the observation, that for many common chemical reactions at room temperature the reaction rate doubles for every 10 degree Celsius increase in temperature.

$$k = C_1 \exp\left(-\frac{E_a}{RT}\right)$$

where C_1 is a constant and E_a the activation energy, R the universal gas constant and T is the temperature. Resulting in the following model the amount of ethyl acetate at time t

$$A_t = \frac{A_0}{1 + A_0\, t\, C_1 \exp\left(-E_a/RT\right)}$$

resulting in a nonlinear model of the form

$$A_t = \frac{\theta_1}{1 + \theta_2\, t \exp\left(-\theta_3/T\right)} + \varepsilon_t$$

with $\theta_1 = A_0$, $\theta_2 = C_1 A_0$, and $\theta_3 = E_a/R$. This is a model in multiple predictors, time and temperature.
Measurements on this model shall allow us to estimate the parameter of the model and estimate A_0, C_1, and E_a.

8.1 Nonlinear Least Squares

In a equivalent approach as to MLRMs, the first goal is to provide estimates for the model parameters. Again the criteria will be to choose the estimates in such a way that the total of the squared distances of the measurements to the estimated mean function $\hat{f}$ is as small as possible.
Mathematically:

$$\min_{\vec{\theta} \in \Theta} S(\vec{\theta}) \quad \text{with} \quad S(\vec{\theta}) = \sum_{i=1}^{n} [y_i - f(\vec{x}_i, \vec{\theta})]^2$$

where the data are pairs of response y_i and predictor value vector $\vec{x}_i$, $1 \le i \le n$.
In order to find the solution $\hat{\theta}$, find the partial derivatives of S with respect to the different parameters and set them to zero, resulting in p normal equations for p parameters.

$$\sum_{i=1}^{n} (y_i - f(\vec{x}_i, \hat{\theta})) \left[\frac{\partial f(\vec{x}_i, \vec{\theta})}{\partial \theta_j}\right]_{\vec{\theta} = \hat{\theta}} = 0 \quad 1 \le j \le p$$

This system of equations is in general a nonlinear system and therefore hard to solve. In many cases no closed - form solution exists and the solution has to be found using numerical (iterative) methods.

Example 8.2.
For the nonlinear example model given above:

$$y = \theta_1 e^{\theta_2 x} + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

then

$$S(\theta_1, \theta_2) = \sum_{i=1}^{n} [y_i - \theta_1 e^{\theta_2 x_i}]^2$$

and the normal equations are:

$$\begin{aligned} \sum_{i=1}^{n} [y_i - \hat{\theta}_1 e^{\hat{\theta}_2 x_i}][e^{\hat{\theta}_2 x_i}] &= 0 \\ \sum_{i=1}^{n} [y_i - \hat{\theta}_1 e^{\hat{\theta}_2 x_i}][\hat{\theta}_1 x_i e^{\theta_2 x_i}] &= 0 \end{aligned}$$

which are equivalent to

$$\begin{aligned} \sum_{i=1}^{n} y_i e^{\hat{\theta}_2 x_i} - \hat{\theta}_1 \sum_{i=1}^{n} e^{2\hat{\theta}_2 x_i} &= 0 \\ \hat{\theta}_1 \sum_{i=1}^{n} y_i x_i e^{\theta_2 x_i} - \hat{\theta}_1^2 \sum_{i=1}^{n} x_i e^{2\hat{\theta}_2 x_i} &= 0 \end{aligned}$$

if $\theta_1 \neq 0$ these are equivalent to

$$\begin{aligned} \sum_{i=1}^{n} y_i e^{\hat{\theta}_2 x_i} - \hat{\theta}_1 \sum_{i=1}^{n} e^{2\hat{\theta}_2 x_i} &= 0 \\ \sum_{i=1}^{n} y_i x_i e^{\hat{\theta}_2 x_i} - \hat{\theta}_1 \sum_{i=1}^{n} x_i e^{2\hat{\theta}_2 x_i} &= 0 \end{aligned}$$

This system is not linear in $\hat{\theta}_1$ and $\hat{\theta}_2$. Each system of normal equations has to be studied in order to find, if the solution is unique, if the iterative methods converge and how to find solutions for the system, etc.
Just finding estimates for the model parameters can be very challenging in nonlinear regression.

8.2 Linearization for finding parameter estimates

Since we know how to find least squares estimators in the linear model, the question arise if we can use this to find solutions for nonlinear models.
One method for finding approximations to the least squares estimators in the nonlinear model is based on the linear terms of the Taylor series of f about a point $\vec{\theta}_0' = (\theta_{10}, \theta_{20}, \ldots, \theta_{p0})$.
The complete Taylor series of f about a point $\vec{\theta}_0' = (\theta_{10}, \theta_{20}, \ldots, \theta_{p0})$ is given by

$$f(\vec{x}, \vec{\theta}) = \sum_{k=0}^{\infty} \sum_{j=1}^{p} \left[\frac{\partial^k f(\vec{x}, \vec{\theta})}{\partial \theta_j^k} \right]_{\vec{\theta}=\vec{\theta}_0} \frac{(\theta_j - \theta_{j0})^k}{k!}$$

Using only the two first terms result in an approximation of f:

$$f(\vec{x},\vec{\theta}) \approx f(\vec{x},\vec{\theta}_0) + \sum_{j=1}^{p} \left[\frac{\partial f(\vec{x},\vec{\theta})}{\partial \theta_j}\right]_{\vec{\theta}=\vec{\theta}_0} (\theta_j - \theta_{j0}) =: g(\vec{x}_i,\vec{\theta},\vec{\theta}_0)$$

The idea for finding the approximate least squares estimates for a given situation is as follows:

1. Start with an initial value $\vec{\theta}_0$.
2. Using the LINEAR function g find the least squares estimate for θ in the MLRM given by this function (the design matrix will have entries from the partial derivatives evaluated at the data points x_i using $\vec{\theta}_0$). Call this solution $\hat{\theta}_1$.
3. Set $\vec{\theta}_1 = \hat{\theta}_1$ (your currently best solution) and repeat step 2 replacing $\vec{\theta}_0$ by $\vec{\theta}_1$ resulting in a new estimate $\hat{\theta}_2$.
4. Continue this process until you do not observe "relevant" changes in R^2_{adj} or the parameter vectors.

The following is a worked out algorithm indicated in the steps above:

1. Set

$$f_i^0 = f(\vec{x}_i,\vec{\theta}_0)$$
$$\beta_j^0 = \theta_j - \theta_{j0}$$
$$Z_{ij}^0 = \left[\frac{\partial f(\vec{x}_i,\vec{\theta})}{\partial \theta_j}\right]_{\vec{\theta}=\vec{\theta}_0}$$

then for $1 \leq i \leq n$

$$y_i - f_i^0 = \sum_{j=1}^{p} \beta_j^0 Z_{ij}^0 + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0,\sigma^2)$$

fit a MLRM. Of course the solution $\hat{\beta}_0$ of this model will only deliver an approximate solution of the NLRM, depending on the start value $\vec{\theta}_0$.

Write above model in matrix notation as

$$\vec{Y} - \vec{f}_0 = \vec{Y}_0 = Z_0\vec{\beta}_0 + \vec{\varepsilon}$$

then an estimate for $\vec{\beta}_0$ is

$$\hat{\beta}_0 = (Z_0'Z_0)^{-1}Z_0\vec{Y}_0 = (Z_0'Z_0)^{-1}Z_0(\vec{Y} - \vec{f}_0).$$

Since

$$\vec{\theta} \approx \vec{\beta}_0 + \vec{\theta}_0$$

therefore an improvement on the initial guess should be given by

$$\hat{\theta}_1 = \hat{\beta}_0 + \hat{\theta}_0$$

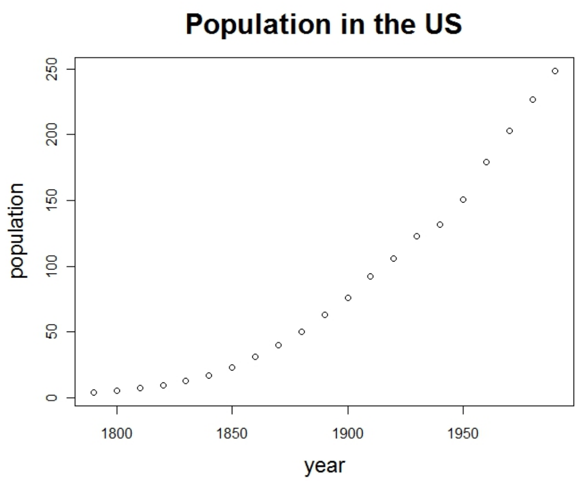
2. We can continue from here and continue to improve $\hat{\theta}_1$ using the same arguments as before.

   In general
$$\hat{\theta}_{k+1} = \hat{\theta}_k + \hat{\beta}_k = \hat{\theta}_k + (Z_k' Z_k)^{-1} Z_k (\vec{Y} - \vec{f}_k)$$
   with
$$f_i^k = f(\vec{x} - i, \vec{\theta}_k)$$
$$Z_{ij}^k = \left[\frac{\partial f(\vec{x}_i, \vec{\theta})}{\partial \theta_j} \right]_{\vec{\theta}=\vec{\theta}_k}$$

3. The process ends, when a convergence criteria is met, like, no relevant changes in the estimates, or the residual sum of squares is not significantly decreasing.

Example 8.3.
For the example we will be using data on the population of the United States (Source: Statistical Abstract of the United States(1994), Bureau of the Census.)

year	population	year	population
1790	3.929	1900	75.995
1800	5.308	1910	91.972
1810	7.240	1920	105.711
1820	9.638	1930	122.775
1830	12.866	1940	131.669
1840	17.069	1950	150.697
1850	23.192	1960	179.323
1860	31.443	1970	203.302
1870	39.818	1980	226.542
1880	50.156	1990	248.710
1890	62.948		



A function for modeling population growth is
$$f(x) = \frac{\beta_1}{1 + e^{\beta_2 + \beta_3 x}}$$

$f(x)$ is the population size at time x, β_1 is the asymptote towards which the population grows, β_2 determines the population size at time 0 (given β_1), and β_3 measures the growth rate of the population.
The function implies the following model equation for individual measurements in the population
$$Y_i = f(x_i) + \varepsilon_i \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2) \quad 1 \leq i \leq n$$

R provides function `nls` for finding estimates for nonlinear regression models. It requires initial values for the parameters, as we have seen that the iterative algorithm depends on it.

The growth curve is S-shaped so one can guess from the scatterplot a value for $\beta_1^0 = 400$, then for $x = 0$ being year 1790

$$3.929 = \frac{400}{1 + e^{\beta_2^0}} \Leftrightarrow \beta_2^0 = \ln\left(\frac{400}{3.929} - 1\right) \approx 4.61$$

To find a good starting value β_3^0, one can use any other data point for example at time $x = 1$

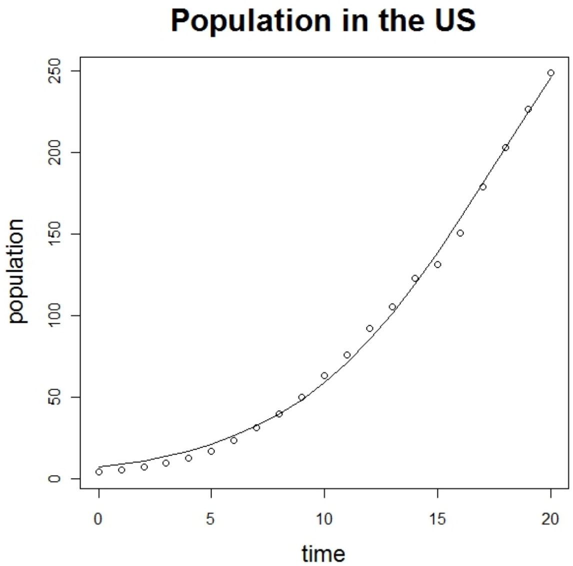
$$5.308 = \frac{400}{1 + e^{4.61+\beta_3^0 * 1}} \Leftrightarrow \beta_3^0 = \ln\left(\frac{400}{5.308} - 1\right) - 4.61 \approx -0.30$$

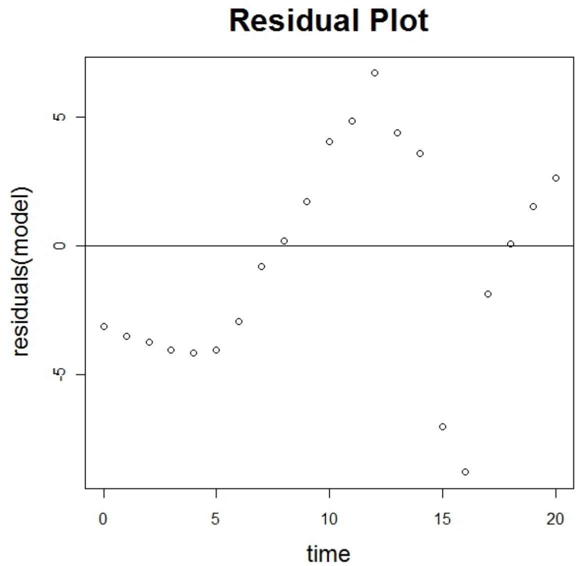
The R code and output is:

```
beta10 <- 400
beta20 <- log(beta10/3.929 -1)
beta30 <- log(beta10/5.308 -1)-beta20

time<-(0:20) # otherwise another parameter has to be included to adjust for the year

model<-nls(pop~beta1/(1+exp(beta2+beta3*time)), start=list(beta1=beta10,beta2=beta20, beta3=b
summary(model)


36737.47 :  400.000000    4.613209   -0.304318
436.1164 :  356.3585914   3.9175587  -0.2337715
368.0317 :  383.3007842   3.9802129  -0.2270488
356.4056 :  388.8121502   3.9903049  -0.2267113
356.4001 :  389.1630093   3.9903011  -0.2266178
356.4001 :  389.1642483   3.9903471  -0.2266204
356.4001 :  389.1656872   3.9903451  -0.2266198

Formula: pop ~ beta1/(1 + exp(beta2 + beta3 * time))

Parameters:
       Estimate Std. Error t value Pr(>|t|)
beta1 389.16569   30.81201   12.63  2.2e-10 ***
beta2   3.99035    0.07032   56.74  < 2e-16 ***
beta3  -0.22662    0.01086  -20.87  4.6e-14 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 4.45 on 18 degrees of freedom

Number of iterations to convergence: 6
Achieved convergence tolerance: 1.016e-06
```

All three parameters are significantly different from 0.

The fitted function and the residual plots show:





The fitted function seems to fit the data well, but the residual plot illustrates that the model is missing some information. Another variable explaining the change in growth rate could help the model. One possible variable could be an indicator of war activities or economic depression.