IOP Conference Series: Materials Science and Engineering

PAPER • **OPEN ACCESS**

A different approach to estimate nonlinear regression model using numerical methods

To cite this article: B Mahaboob *et al* 2017 *IOP Conf. Ser.: Mater. Sci. Eng.* **263** 042122

IOP Conf. Series: Materials Science and Engineering **263** (2017) 042122 doi:10.1088/1757-899X/263/4/042122

A different approach to estimate nonlinear regression model using numerical methods

B Mahaboob[1], B Venkateswarlu[2], G Mokeshrayalu[2] and P Balasiddamuni[3]
[1]Department of Mathematics, Swetha Institute of Technology, and Science, Tirupati, India
[2]Department of Mathematics, School of Advanced Sciences, VIT University, Vellore-632014, India
[3]Department of Statistics, S.V.University, Tirupati, India

E-mail: venkatesh.reddy@vit.ac.in

Abstract. This research paper concerns with the computational methods namely the Gauss-Newton method, Gradient algorithm methods (Newton-Raphson method, Steepest Descent or Steepest Ascent algorithm method, the Method of Scoring, the Method of Quadratic Hill-Climbing) based on numerical analysis to estimate parameters of nonlinear regression model in a very different way. Principles of matrix calculus have been used to discuss the Gradient-Algorithm methods. Yonathan Bard [1] discussed a comparison of gradient methods for the solution of nonlinear parameter estimation problems. However this article discusses an analytical approach to the gradient algorithm methods in a different way. This paper describes a new iterative technique namely Gauss-Newton method which differs from the iterative technique proposed by Gorden K .Smyth [2]. Hans Georg Bock et.al [10] proposed numerical methods for parameter estimation in DAE's (Differential algebraic equation). Isabel Reis Dos Santos et al [11], Introduced weighted least squares procedure for estimating the unknown parameters of a nonlinear regression metamodel. For large-scale non smooth convex minimization the Hager and Zhang (HZ) conjugate gradient Method and the modified HZ (MHZ) method were presented by Gonglin Yuan et al [12].

1. Introduction

A great deal of research in mathematical modelling has been directed to nonlinear modelling and establishing functional relationships among different variables. Nonlinear models have a wide number of applications in physical, biological and social sciences, business, engineering, economics and management sciences. Now-a-days nonlinear model building in new and very fascinating field of research in applied mathematical sciences. In Mathematics or in any other scientific discipline a research worker is certainly facing the problem of formulation of a nonlinear model. A large number of nonlinear models have been specified in the literature and successfully applied to different situations in the real world relating to several research problems in the various fields of applied mathematics. However there are a large number of situations which have not yet been nonlinearly modelled because of the situations may be complex or they are mathematically or statistically intractable. In the present research study gives some new methods of solving nonlinear normal equations in order to estimate parameters of nonlinear regression models.Goldfeld and Quandt [3] have provided various computational methods based on numerical analysis to estimate the parameters of nonlinear regression model. Suppose that the asymptotic properties are used and the errors are normally distributed or the error distribution is unknown. The least squares estimators can be obtained in this case for the parameters of the nonlinear regression model as good choice. But, there are some problems involved in getting these estimators. First, nonlinear normal equations to be obtained by minimizing residual sum of squares cannot usually be solved analytically and second, may be multiple solutions not all corresponding to the global minimum of the residual sum of their squares function. The solutions of the nonlinear normal equations can be obtained by using computational methods.

2. The Gauss – newton method

Consider the linear pseudo model for nonlinear regression model $Y = f(X, \beta) + \varepsilon$ by using linear approximation at $\beta = \beta^*$ as

$$Y^* = Z(\beta^*)\beta + \varepsilon \tag{2.1}$$

Where $Y^* = Y - f(X, \beta^*) + Z(\beta^*)\beta^*$ (2.2)

Generally, linear approximation will only be good close to the point where the derivative is evaluated. Since, the true parameter vector is unknown such that the nonlinear regression model cannot be approximated by linear model close to β^* one may use some other vector say β_1^*, for β^*.

Now, the pseudo model can be written as

$$Y_1^* = Z(\beta_1^*)\beta + \varepsilon \tag{2.3}$$

The linear least squares estimator of β is given by

$$\beta_2^* = \left[Z(\beta_1^*)' Z(\beta_1^*)\right]^{-1}\left[Z(\beta_1^*)' Y_1^*\right] \tag{2.4}$$

Assuming that β_2^* is closer than β_1^* to the solution of the nonlinear normal equations, then one may compute the next estimate replacing β_1^* by β_2^* in (2.3) and obtain

$$\beta_3^* = \left[Z(\beta_2^*)' Z(\beta_2^*)\right]^{-1}\left[Z(\beta_2^*)' Y_2^*\right] \tag{2.5}$$

If β_3^* differs from β_2^*, one may replace β_2^* by β_3^* in (2.5) to obtain β_4^*, and so on.

This method is an iterative process. In the n^{th} iteration, one may compute

$$\beta_{n+1} = \left[Z(\beta_n^*)' Z(\beta_n^*)\right]^{-1}\left[Z(\beta_n^*)' Y_n^*\right] \tag{2.6}$$

The iteration stops if $\beta_{n+1} = \beta_n$

Using (2.2), one may write (2.6) as

$$\hat{\beta}_{n+1} = \left[Z(\beta_n^*)' Z(\beta_n^*)\right]^{-1} Z(\beta_n^*)'\left[Y - f(X, \beta_n^*) + Z(\beta_n^*)\beta_n^*\right] \tag{2.7}$$

$$\left[Z(\beta_n^*)' Z(\beta_n^*)\right]\beta_{n+1}^* = Z(\beta_n^*)'\left[Y - f(X, \beta_n^*)\right] + \left[Z(\beta_n^*)' Z(\beta_n^*)\right]\beta_n^* \tag{2.8}$$

If $\beta_{n+1}^* = \beta_n^*$ this implies that

$$-2Z(\beta_n^*)'\left[Y - f(X, \beta_n^*)\right] = 0 \tag{2.9}$$

In practice, one may continue the iterative process until β_{n+1}^* almost equals β_n^* in the two consecutive cycles of process. This method of solving the nonlinear normal equations is known as the

"Gauss Method" or Newton – Gauss Algorithm. The method of choosing β_1^* has been suggested by Chow [4] and Chow and Fair [5].

3. Gradient algorithm methods

Newton-Raphson method

The Newton-Raphson method can be considered a generalization of the Gauss-Newton method for estimating nonlinear parameters. Suppose that L (β) be the likelihood function which has continuous first and second order derivatives with respect to the parametric vector β of p – parameters. For the maximization of L (β) one can set its first order derivative equal to zero.

i.e. $$\frac{\partial L(\beta)}{\partial \beta} = L'(\beta) = 0 \tag{3.1}$$

In general, $L'(\beta)$ is known as the gradient. To maximize the likelihood function $L(\beta)$, many algorithms are based on the gradient $L'(\beta)$ and the algorithms are called the gradient algorithms. Generally an algorithm specifies how one should move from a point β^0 to the next point β^1. A gradient algorithm takes the form

$$\beta^{(1)} = \beta^0 + k^0 H^0 L'(\beta^0) \tag{3.2}$$

Where, $L'(\beta^0)$ is the gradient evaluated at β^0; k^0 is a scalar; H^o is a matrix to be specified; k^0 gives the step size.

and $H^0 L'(\beta^0)$ is the "Search Direction".The Newton-Raphson method specifies

$$H^0 = -\left[\frac{\partial L'(\beta^0)}{\partial \beta^{(1)}}\right]^{-1} \tag{3.3}$$

In other words, the set of nonlinear equations $L'(\beta) = 0$ can be solved by using the Newton-Raphson method, in which one may start with an initial value β^0 and the Newton-Raphson method linearizes $L'(\beta)$ about $\beta = \beta^0$ as

$$L'(\beta) = L'(\beta^0) + \left[\frac{\partial L'(\beta)}{\partial \beta'}\right]_{\beta^0} (\beta - \beta^0) = 0 \tag{3.4}$$

The solution of this linear equation gives

$$\beta^{(1)} = \beta^0 - \left[\frac{\partial L'(\beta)}{\partial \beta'}\right]_{\beta^0}^{-1} L'(\beta^0) \tag{3.5}$$

Equations (3.1) and (3.1)

$$\Rightarrow H^0 = -\left[\frac{\partial L'(\beta^0)}{\partial \beta^{(1)}}\right]^{-1}$$

Thus, if $\frac{\partial L'(\beta)}{\partial \beta}$ is linear or L (β) is quadratic, the Newton-Raphson method using k=1 converges in one iteration. That is, starting from any β^0, the solution $\beta^{(1)}$ in (3.5) gives the maximum of L(β).

Steepest descent or steepest ascent algorithm method

Consider the general form of gradient algorithm to maximize the likelihood function L (β) for the first iteration as

$$\beta^{(1)} = \beta^0 + k^0 \ H^0 \ L'(\beta^0) \tag{3.6}$$

Where $L'\left(\beta^0\right) = \left[\frac{\partial L\left(\beta\right)}{\partial \beta}\right]_{\beta=\beta^0}$ (3.7)

k^0 is a step size to be chosen as a scalar, H^0 is Hermitian matrix to be specified andβ^0 is initial value of β, $H^0 L'\left(\beta^0\right)$ is the search direction at $\beta = \beta^0$, $L'\left(\beta^0\right)$ is the gradient evaluated at β^0.The method of Steepest Descent specifies H^0 = - I and the method of Steepest Ascent specifies H^0 = I in all the iterations. One may choose k^0 to approximate L (β) by a quadratic function and find k^0 to maximize L ($\beta^{(1)}$) .

Consider the approximate quadratic function form of $L'\left(\beta^{(1)}\right)$ as

$$L\left(\beta^{(1)}\right) = L\left(\beta^0\right) + \left[\frac{\partial L\left(\beta^0\right)}{\partial \beta'}\right]\left(\beta^{(1)} - \beta^0\right) + \frac{1}{2}\left(\beta^{(1)} - \beta^0\right)'\left[\frac{\partial^2 L\left(\beta^0\right)}{\partial \beta \, \partial \beta'}\right]\left(\beta^{(1)} - \beta^0\right) \tag{3.8}$$

By substituting $k \ L'\left(\beta^0\right)$ for $\left(\beta^{(1)} - \beta^0\right)$ and minimizing with respect to k gives

$$k^0 = -\left[L'\left(\beta^0\right)\right]'\left[L'\left(\beta^0\right)\right]\left\{\left[L'\left(\beta^0\right)\right]'\left[\frac{\partial L'\left(\beta^0\right)}{\partial \beta}\right]\left[L'\left(\beta^0\right)\right]\right\}^{-1} \tag{3.9}$$

In practice, this method may not be useful because, it may converge slowly.

This method can be applied if it is combined with other algorithm methods such as Gauss-Newton method, Newton – Raphson method etc.

The method of scoring

Consider the Gradient Algorithm form as

$$\beta^{(1)} = \beta^0 + k^0 \ H^0 \ L'\left(\beta^0\right) \tag{3.10}$$

Here, $L'\left(\beta^0\right) = \left[\frac{\partial L\left(\beta\right)}{\partial \beta}\right]_{\beta=\beta^0}$

In the maximum likelihood method of estimation, the quantity $\left[\frac{\partial \operatorname{Log} L\left(\beta\right)}{\partial \beta}\right]$ is sometimes known as "Efficient score for β". The maximum likelihood equation to get maximum likelihood estimator $\hat{\beta}$ for β is given by

$$\left[\frac{\partial \operatorname{Log} L\left(\beta\right)}{\partial \beta}\right]_{\beta=\hat{\beta}} = 0 \tag{3.11}$$

That is, $\hat{\beta}$ is the solution of the maximum likelihood equation, for which the efficient score $\left[\frac{\partial \operatorname{Log} L\left(\beta\right)}{\partial \beta}\right]_{\beta=\hat{\beta}}$ vanishes. Very often, the maximum likelihood equation is nonlinear and it can be solved by using Gradient Algorithm under Iterative method. The method of scoring specifies $H^0 = \left[I\left(\beta^0\right)\right]^{-1}$

Where $I\left(\beta^0\right) = -E\left[\frac{\partial^2 \operatorname{Log} L(\beta)}{\partial\beta\,\partial\beta'}\right]_{\beta=\beta^0}$ (3.12)

As in the Newton-Raphson method using $k^0=1$, the Gradient Algorithm in the first iteration,

$$\beta^{(1)} = \beta^0 + \left[I\left(\beta^0\right)\right]^{-1}\left[\frac{\partial \log L\left(\beta^0\right)}{\partial\beta^0}\right] \quad (3.13)$$

gives the maximum of L(β). In the second iteration, one may obtain,

$$\beta^{(2)} = \beta^{(1)} + \left[I\left(\beta^{(1)}\right)\right]^{-1}\left[\frac{\partial \log L\left(\beta^{(1)}\right)}{\partial\beta^{(1)}}\right] \quad (3.14)$$

Iterative process continues till to get convergence.

The method of quadratic hill-climbing

Goldfield and Quandt and Trotter [6] have proposed this method as a modification of the Newton-Raphson method. If β^0 is far from the maximizing value, then the second order partial derivative may not be negative definite. To ensure the negative definiteness of $-H^0$, the method of quadratic hill climbing uses for $-H^0$ in Gradient Algorithm as

$$\left\{\left[\frac{\partial L'(\beta)}{\partial\beta'}\right]_{\beta=\beta^0} - \alpha I\right\} \quad (3.15)$$

Where the scalar α is chosen to maximize L (β) in a spherical region centered at β^0.

In other words, if L (β) is assumed to be quadratic in the spherical region, then it is bounded by

$$\left(\beta-\beta^0\right)'\left(\beta-\beta^0\right) = r_\alpha \quad (3.16)$$

The characteristic roots of the matrix $\left[\frac{\partial L'(\beta)}{\partial\beta'}\right]_{\beta=\beta^0}$ are to be computed for this method.

Many other Gradient Algorithms are available in the literature. For instance, Powell developed the conjugate - gradient method, which evaluates the likelihood function $L(\beta)$ to be maximized along mutually conjugate directions, beginning at a point β^0. Goldfeld and Quandt [7] and Quandt [8] have discussed various Gradient Algorithms and developed a useful computer package GQOPT which includes several algorithms for optimization.

4. Some important nonlinear regression growth models

Following are some important nonlinear regression growth models given in the literature:

(i) Nonlinear compound growth model :

$$Y_t = \beta_0\beta_1^t\varepsilon \quad (4.1)$$

(ii) Nonlinear second degree parabolic growth model :

$$Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \varepsilon \quad (4.2)$$

(iii) Nonlinear m^{th} degree polynomial growth model:

$$Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + ... + \beta_m t^m + \varepsilon \quad (4.3)$$

(iv) Nonlinear exponential growth model :

$$Y_t = \beta_0 e^{\beta_1^t}\varepsilon \quad (4.4)$$

(v) Nonlinear second degree polynomial growth model :

$$Y_t = \beta_0 \beta_1^{t} \beta_2^{t^2} \varepsilon \tag{4.5}$$

(vi) Nonlinear modified exponential growth model :

$$Y_t = \beta_0 + \beta_1 \beta_2^{t} \varepsilon \tag{4.6}$$

(vii) Nonlinear Gompertz growth model :

$$Y_t = \beta_0 \beta_1^{\beta_2^{t}} \varepsilon \tag{4.7}$$

(viii) Nonlinear logistic growth models :

$$\left.\begin{aligned} &(a)\ Y_t = \frac{k}{1+e^{\beta_0+\beta_1 t+\varepsilon}} \\ &(b)\ Y_t = \frac{k}{1+\beta_0 e^{-\beta_1 t+\varepsilon}} \end{aligned}\right\} \text{Where} \quad \beta_1 < 0 \tag{4.8}$$

(ix) Nonlinear Log – Logistic growth model :

$$Y_t = \frac{k}{1+\beta_0 e^{-\beta_1 \mathrm{Log}t+\varepsilon}} \tag{4.9}$$

(x) Nonlinear modified Gompertz growth model :

$$Y_t = \beta_0 \exp\left[-\beta_1 \exp\left(-\beta_2 t\right)\right]\varepsilon \tag{4.10}$$

(xi) Nonlinear negative exponential growth model :

$$Y_t = \beta_0 \left[1 - e^{-\beta_1 t}\right]\varepsilon \tag{4.11}$$

(xii) Nonlinear negative Gompertz growth model:

$$Y_t = \beta_0 \left[1 - \beta_1 e^{-\beta_2 t}\right]\varepsilon \tag{4.12}$$

(xiii) Nonlinear Richard's growth model :

$$Y_t = \frac{k}{1+\beta_0\left[\exp\left(-\beta_1 t\right)\right]}\varepsilon \tag{4.13}$$

(xiv) Nonlinear Von Bertalanffy growth model :

$$Y_t = \left[\beta_0^{1-\beta_3} - \beta_1 \exp\left(-\beta_2 t\right)\right]^{\frac{1}{1-\beta_3}} \tag{4.14}$$

(xv) Nonlinear Wei-bull growth model :

$$Y_t = \beta_0 - \beta_1 \exp\left(-\beta_2 t^{\beta_3}\right) + \varepsilon \tag{4.15}$$

(xvi) Nonlinear Schnute growth model :

$$Y_t = \beta_0 + \beta_1 \exp\left(\beta_2 t\right)^{\beta_3} + \varepsilon \tag{4.16}$$

(xvii) Nonlinear Morgan – Mcreer – Flodin growth model :

(xviii)

$$\left.\begin{aligned} &a)\ Y_t = \beta_0 - \left[\frac{\beta_0 - \beta_1}{\left(1+\left(\beta_2^{t}\right)\right)^{\beta_0}\varepsilon}\right] \\ &b)\ Y_t = \left[\frac{\beta_0\beta_1 + \beta_2 t^{\beta_3}}{\beta_1 + t^{\beta_3}}\right]\varepsilon \end{aligned}\right\} \tag{4.17}$$

(xix) Nonlinear Chapman – Richard's growth model :

$$Y_t = \beta_0 \left[1 - \beta_1 \exp\left(-\beta_2 t\right)\right]^{\frac{1}{1-\beta_3}} \varepsilon \quad (4.18)$$

(xx) Nonlinear Standard growth model :

$$Y_t = \beta_0 \left[1 + \beta_3 e^{-(\beta_1 + \beta_2 t)}\right]^{\beta_4} \varepsilon \quad (4.19)$$

5. Conclusions

Generally optimal estimators for the parameters of nonlinear model that is intrinsically linear can be obtained by applying Ordinary Least Squares (OLS) estimation method to the transformed model. The OLS estimation fails to give estimators for the parameters of nonlinear model that is intrinsically nonlinear. However iterative OLS estimation method can be applied to estimate parameters of this model. In the above research work some new numerical techniques for estimating parameters has been discussed analytically using principles in matrix calculus and some important nonlinear regression growth models available in the literature are mentioned.

References

[1] Yonathan Bard1970 SIAM *Journal on Numerical Analysis***7**(1) 157-186
[2] Gordon K and Smyth *Encyclopedia of Environ metrics* **3**1405-1411
[3] Goldfeld S M and Quandt RE1972 North-Holland Publishing Co Amsterdam
[4] Chow G C *1973Review of Economics and Statistics* **55** 104- 109
[5] Chow G C and RC Fair 1973 *Annals of Economics and Social Measurement*
[6] Goldfield S M RE Quandt and H F Trotter 1966 *Econometric* a**34**541-551
[7] Goldfeld S M and Quandt R E 1972 North-Holland Publishing Co Amsterdam
[8] Quandt R E ZGiriliches and M D Intriligator1983eds Handbook of EconometricsAmsterdam:North-Holland**1** 699-764
[9] Powell M D 1964 *Computer journal* **7** 115-162
[10] Hans Georg Bock Ekaterina Kostina and Johannes P Schloder 2007 GAMM-Mitt**30**(2) 376-408
[11] M Isabel Reis Dos Santos and Acacio M O Forta Nova 2007 *Communications in Statistics-Stimulation and Computation* **36**123-137
[12] Yuan G Shang Zand Liu W 2016 *PLOS ONE* **11** 10