Numerical techniques of nonlinear regression model estimation

Cite as: AIP Conference Proceedings **2177**, 020082 (2019); https://doi.org/10.1063/1.5135257
Published Online: 04 December 2019

Ranadheer Donthi, J. Peter Praveen, S. Vijay Prasad, B. Mahaboob, and B. Venkateswarlu

Numerical Techniques of Nonlinear Regression Model Estimation

Ranadheer Donthi[1, b)], J. Peter Praveen[2], S. Vijay Prasad[2], B.Mahaboob[2], B.Venkateswarlu[3, a)]

[1]*St. Martin's Engineering College, Dhulapally, Kompally, Hyderabad, Telangana, India 500100*
[2]*Department of Mathematics, Koneru Lakshmaih Education Foundation, Vaddeswaram, Guntur Dist.AP-522502-India*
[3]*Department of Mathematics, Vellore Institute of Technology, Vellore, Tamilnadu*

[a)]Corresponding author: venkatesh.reddy@vit.ac.in
[b)] ranadheer.donthi@gmail.com

Abstract. The literature on numerical methods for fitting nonlinear regression model has grown enormously in the fast five decades. An important phase in nonlinear regression problems is the exploration of the relation between the independent and dependent variables. A largely unexplored area of research in nonlinear regression models concerns the finite sample properties of nonlinear parameters. The main object of this research study is to pro- pose some nonlinear methods of estimation of nonlinear regression models, namely Newton- Raphson method, Gauss-Newton method, Method of scoring, Quadratic Hill-Climbing and Conjugate Gradient methods. In 2005, G.E. Hovland et al (see [5]). In his research article, presented a parameter estimation of physical time-varying parameters for combined-cycle power plant models. B. Mahaboob et al. (see [6]), in their research paper, proposed some computational methods based on numerical analysis to estimate the parameters of nonlinear regression model. S.J. Juliear et al. (see [7]), in their research paper, developed the method of unscented transformation (UT) to propagate mean and covariance information through nonlinear transformations.

INTRODUCTION

A model is treated as a mathematical description of a physical or chemical or biological state. Most of the models occur in Applied Mathematical and Statistical Sciences are nonlinear in nature. One of the most important aspects in the literature of Theoretical and Applied Mathematics is the inferential procedures i.e estimating the parameters and testing the hypothesis about the parameters of nonlinear statistical model. A perfect model can have too many parameters of nonlinear statistical models. A perfect model can have too many parameters to be useful. Nonlinear regression models have been intensively studied in the last fifty years. The main contributions in the field of nonlinear regression models have been made by Gallant, Rosi and Tauchen (1933), Levenberg (1944), Hartley(1961), Jenrich (1969), Goldfeld and Quandt (1970), Biggs(1971), Ross(1971), Chambers(1973), Gallant (1975a,1975b), Bates and Watts (1980,2008), Dennis, Gay and Welsc (1981), Hiebert (1981), McCullgh (1983), Ratkowsky (1983), Dennis and Schnabel (1983), Cordiero and Paula (1989), Taylor and Uhlig (1990), Cameron Windmeijer (1997), Ord, Koehler and Sny- der (1997), Davidson and Mackinnon (1999), Popli (2000), Fox (2002), Smyth (2002), Davidian and Giltinan (2003), Vasilyer (2008), Fox and Wiesberg (2010), Potocky and Stehlik(2018), Grafarend and Awange (2012), Frost (2013) and others.

Recent Trends in Pure and Applied Mathematics
AIP Conf. Proc. 2177, 020082-1–020082-6; https://doi.org/10.1063/1.5135257
Published by AIP Publishing. 978-0-7354-1924-7/$30.00

NEWTON RAPSON METHOD

Let a nonlinear regression model expressed in matrix from be $Y = f(X, \alpha) + \varepsilon$ and $L(\alpha)$ be the LE possessing continuous first and second order derivatives with respect to vector α of r-elements. For the optimization of $L(\alpha)$ one can put

$$\frac{\partial L(\alpha)}{\partial \alpha} = 0. \tag{1}$$

At any point α, the gradient of first derivatives $L^1(\alpha)$ gives more information at which one can reach maximum. Gradient algorithms are applied to maximize $L(\alpha)$.

Using this one can get

$$\alpha^{(1)} = \alpha^{(0)} + M^0 N^0 L^1(\alpha^{(0)}). \tag{2}$$

Here $L^1(\alpha^{(0)})$ is gradient evaluated at $(\alpha^{(0)})$.

M^0 is a scalar and gives step size.

N^0 is a matrix.

$N^0 L^1(\alpha^{(0)})$ is direction of search.

The Newton Rapson method specifies

$$N^0 = -\left(\frac{\partial L^1(\alpha)}{\partial \alpha}\right)^1_{\alpha=\alpha^{(0)}}. \tag{3}$$

So N^0 is the inverse of Hessian Matrix. If $L^1(\alpha)$ is linear the Newton- Rapson method using M=1 converges in just one iterative step.

That is

$$\alpha^{(1)} = \alpha^{(0)} - \left(\frac{\partial L^1(\alpha)}{\partial \alpha}\right)_{\alpha=\alpha^{(0)}} \left(L^1(\alpha^{(0)})\right). \tag{4}$$

Gives the maximum of $L(\alpha)$.

If $L(\alpha)$ is not quadratic one can chose the step size $M^0 \neq 1$.

STEEPEST DESCENT/ASCENT METHOD

This suggests $N^0 = I.$ One can select step size M^0 is to approximate $L(\alpha)$ by second order polynomial function and find M^0 to optimize $L^1(\alpha)$.

Suppose

$$L(\alpha^{(1)}) = L(\alpha^{(0)}) + \frac{\partial L(\alpha^{(0)})}{\partial \alpha}(\alpha^{(1)} - \alpha^{(0)}) + \frac{1}{2}(\alpha^{(1)} - \alpha^{(0)})^{(1)} \frac{\partial^2 L(\alpha^{(0)})}{\partial \alpha \partial \alpha^1}(\alpha^{(1)} - \alpha^{(0)}). \tag{5}$$

Put $(\alpha^{(1)} - \alpha^{(0)}) = ML^{(1)}(\alpha^{(1)})$

Now (5) becomes

$$L(\alpha^{(1)}) = L(\alpha^{(0)}) + M[L^1(\alpha^{(0)})]^1[L^1(\alpha^{(0)})] + \frac{1}{2}M^2[L^1(\alpha^{(0)})]^1 \frac{\partial L^1(\alpha^{(0)})}{\partial \alpha}[L^1(\alpha^{(0)})]. \quad (6)$$

Minimization of (6) with respect to M implies

$$M^0 = -[L^1(\alpha^{(0)})]^1[L^1(\alpha^{(0)})] + \left(\left([L^1(\alpha^{(0)})]^1 \frac{\partial L^1(\alpha^{(0)})}{\partial \alpha}\right)[L^1(\alpha^{(0)})]\right)^{-1}. \quad (7)$$

Here M^0 is tedious to calculate as it is required the matrix $\frac{\partial L^1(\alpha^{(0)})}{\partial \alpha}$. This method seems to be easy but cannot possess practical use as it is having slow convergence. Nevertheless this method is very useful if it is mixed with some algorithms eg Gauss Newton Method.

GAUSS NEWTON METHOD

This is an approximation to Newton Rapson method and in that

(i) The matrix second derivatives of L are calculated.

(ii) The second order derivatives of $f(X,\alpha)$ which specifies the nonlinear model are not considered. Take the nonlinear model $Y = f(X,\alpha) + \varepsilon$ and the objective function is of the form

$$R(\alpha) = (Y - f(X,\alpha))^1 \psi (Y - f(X,\alpha)). \quad (8)$$

With $\psi = \left((X_{ab})\right)$ *for* $b = 1, 2, ...m.$

The Hessian of $R(\alpha)$ is

$$N(\alpha) = 2\left(A(\alpha)\right)^1 \psi A(\alpha) - 2\sum X_{ab}(Y_a - f(X_a,\alpha))\left(\frac{\partial^2 f(X_a,\alpha)}{\partial \alpha \partial \alpha^1}\right). \quad (9)$$

But, as the average of $\varepsilon_i = f(X_a,\alpha^*)$ is taken as 0, these error terms should be small, at least close to minimum of $R(\alpha)$ if the variance is small. Hence the first term of the RHS of (9) is taken as approximation for $N(\alpha)$.

Hence one can get

$$\left(N(\alpha)\right)^{-1} = \left((A(\alpha_n))^1 \psi (A(\alpha_n))\right). \quad (10)$$

This process yields

$$\alpha_{m+1} = \alpha_m + \left((A(\alpha_m))^1 A(\alpha_m))^{-1} A(\alpha_m)\right)^1 (Y - f(X,\alpha_m)). \quad (11)$$

$$\alpha_{m+1} = \left((A(\alpha_m))^1 A(\alpha_m))^{-1} A(\alpha_m)\right)^1 (Y - f(X,\alpha_m) + A(\alpha_m)\alpha_m). \quad (12)$$

The least squares estimator for the model now becomes

$$Y - f(X,\alpha_m) + A(\alpha_m)\alpha_m = A(\alpha_m)\alpha + \varepsilon. \quad (13)$$

This linear psendo model at α_m. The above procedure depicts that the Gauss- Newton algorithm is seen as a sequence of linear regressions and in this method at each step one can evaluate the LSE for a linear approximation of the nonlinear regression method.

MEHOD OF SCORING

This method is slight changing in Newton Rapson method and can be applied to MLES of the parameters. In matrix calculus the necessary condition for $\log L(\alpha)$ to have optimum at $\alpha = \hat{\alpha}$ is

$$\left(\frac{\partial \log L(\alpha)}{\partial \alpha}\right) = 0. \tag{14}$$

Hence $L(\alpha)$ is the likelihood mapping and the first order partial derivative $\left(\frac{\partial \log L(\alpha)}{\partial \alpha}\right)$ is known as the efficient score for α and is denoted by $B(\alpha)$. Some times (14) can be solved easily. Nevertheless it is nonlinear and open can apply iterative method on it and for this reason one of the gradient methods are to be used. Using the method of scoring, for one parameter stage

$\left(\frac{\partial \log L(\alpha)}{\partial \alpha}\right)$ is expanded by Taylors series at . By neglecting the terms from second order one can have

$$\left(\frac{\partial \log L(\alpha)}{\partial \alpha}\right) = \left(\frac{\partial \log L(\alpha)}{\partial \alpha^0}\right) + (\alpha - \alpha^0)\left(\frac{\partial^2 \log L(\alpha)}{\partial \alpha^{0^2}}\right). \tag{15}$$

put $k(\alpha^0) = \left(\frac{\partial^2 \log L(\alpha)}{\partial \alpha^{0^2}}\right)$ in large samples and get

$$\left(\frac{\partial \log L(\alpha)}{\partial \alpha}\right) \approx B(\alpha^0) - (\alpha - \alpha^0)K(\alpha^0)$$

This with (14) gives

$$\alpha = \alpha^0 + \frac{B(\alpha^0)}{K(\alpha^0)}. \tag{16}$$

$$\alpha^1 = \alpha^0 + \frac{B(\alpha^0)}{K(\alpha^0)}. \tag{17}$$

In matrix symbols

$$\alpha^1 = \alpha^0 + (K(\alpha^0))^{-1} B(\alpha^0). \tag{18}$$

Where $K(\alpha^0) = -E\left(\frac{\partial^2 \log L(\alpha)}{\partial \alpha \partial \alpha^1}\right)_{\alpha=\alpha^0}$, $B(\alpha^0) = -E\left(\frac{\partial \log L(\alpha)}{\partial \alpha}\right)_{\alpha=\alpha^0}$.

The second iteration is

$$\alpha^2 = \alpha^1 + (K(\alpha^1))^{-1} B(\alpha^1). \tag{19}$$

This is called till one can reach the convergence.

QUADRATIC HILL CLIMING METHOD

The matrix $\left(\frac{\partial L^1(\alpha)}{\partial \alpha^1}\right)_{\alpha=\alpha^0}$ of second partial derivatives may not be negative definite when α^0 is from the maximizing value. By choosing a small step in Newton-Rapson direction leads one downhill rather than uphill. The method of quadratic climbing uses for $-N^0$ in the gradient algorithm in order to ensure the negative definiteness of $-N^0$

$$\alpha^1 = \alpha^0 + M^0 N^0 L^1(\alpha^0)$$

$$-N^0 = \left(\frac{\partial L^1(\alpha^0)}{\partial \alpha^1} - \gamma K\right)^{-1}. \tag{20}$$

Here the scalar γ is selected to maximize $L(\alpha)$ is spherical region whose center is α^0 that is bounded by $(\alpha-\alpha^0)^1(\alpha-\alpha^0) = S_\gamma$ under the assumption that $L(\alpha)$ is quadratic in that region. This methods needs the evaluation of Eigen roots of the matrix $\frac{\partial L^1(\alpha^0)}{\partial \alpha^1}$.

CONJUGATE GRADIENT METHOD

This method does not want the usage of first order differential coefficients. It evaluate the function $L(\alpha)$ to be maximized along mutually conjugate directions beginning at the point α^0.

For the quadratic function $L(\alpha) = \alpha^1 D\phi + \gamma^1 \alpha + d$, two directions e_1 and e2 are conjugate if $e_1^1 D e_2 = 0.$

At A=I_l, the 'l' columns of the unit matrix I_l, are conjugate direction vectors.

Let $e_1^0, e_2^0, \ldots e_l^0$ be linearly independent direction vectors.

Beginning with α^0 , one can investigate along the directions $e_a^0 (a = 1,2..l)$ sequentially, each time going along one direction e_a^0. One can start by investigating along e_1^0, i.e by selecting a scalar $\eta_1 \; to \; Max_{\eta_1} L(\alpha^0 e_1^0).$

Having chosen $\hat{\eta}_1$ one can chooses a scalar $\eta_2 \; to \; Max_{\eta 2} L(\alpha^0 + \eta_1 \hat{e}_1^0 \eta_2 \, e_2^0).$ and so on.

Having searched along all 'l' directions, let

$$\hat{\alpha}^0 = \alpha^0 + \sum_{a=1}^{l} \hat{\eta}_a e_a^0 = \alpha^0 + \Delta. \tag{21}$$

The next step is to $Max \, L(\hat{\alpha}^0 + M\Delta)$

Put $\alpha^1 = (\hat{\alpha}^0 + M\Delta)$.

To begin, let e_a^0 be coordinate directions i.e the 'l' column vector of the identity matrix I_l. The directions for the next iterations are $e_1^0 = e_2^0, e_2^1 = e_3^0 = \ldots\ldots e_l^1 = \Delta.$

In the second iteration one can consider $e_a^1 (a = 1,2..l)$ as $e_a^0 (a = 1,2..l)$ in the iteration and so on.

CONCLUSION

In the above research article some numerical techniques for estimating the parameters of nonlinear regression models Newton-Rapson method, Gauss- Newton method, Method of Scoring, Quadratic Hill- Climbing method and Conjugate Gradient methods have been proposed using the principles of MATRIX CALCULUS. In the context of future research one can extend these ideas to frame some numerical techniques namely, Nonlinear least squares estimation method, Linear approximation method and Maximum Likelihood estimation method using the principles of Multi Variable Calculus.

REFERENCES

1. Oral Gpps Jr, Western J. Agriculture, Economics, 50-63(1983).
2. RajaManish, Prude Uni., West Lafayeffe, USA. (2014).
3. S. Neal, , IEE digital explore, IEE Digital Library., 13, (1968).
4. Thomas Schon,, Dept. Electrical Eng., Sweden. (2003).
5. G.E. Hovland et.al, Science Direct, Elsevier 13, 1341-1355(2005).
6. B. Mahaboob et.al, IOP Conference Series: Material Sci. Eng. 263(2017).
7. S. J. Julier, J.K. Uhimann, Proceedings IEE., 92 (2004).
8. Ratowsky. D.A, Willey Publications., 2, (1990).
9. Ross, G.J.S, J. Appl. Stat. 17, 299-306(1990).
10 Popli, G.K, J. Quantitative Economics. 16, 31-36(2000).