Journal of the Nigerian Association of Mathematical Physics
***Volume 65, (October 2022– August 2023 Issue*), pp173 – 178**
© J. of NAMP

NONLINEAR REGRESSION PARAMETER ESTIMATES USING GENETIC ALGORITHMS

Onoghojobi B[(1)], Olewuezi N. P[(2)] and Omojarabi O[(3)]

[(1) & (3)]Department of Statistics, Federal University Lokoja, Nigeria
[(2)]Federal University of Technology, Owerri

Abstract

Deterministic algorithm such as Gauss Newton and Levenberg - Marquadt are still well established practice for obtaining optimal estimates in nonlinear regression. These methods however, have certain pitfalls of multiple local optimal, non-invertibility, differentiability that results to misleading estimates. Under these circumstances, this study is aimed at using optimization techniques in obtaining optimal estimates of complex nonlinear regression model. We investigated the effectiveness and simplicity of particle swarm optimization and genetic algorithm on five (5) test-bed problems obtained from the National Institute of Standards and Technology (NIST) website. R codes were developed for each model. Each algorithm was tried ten (10) times for each model for at least 100 iterations. The results obtained were displayed on tables and graphs. Particle Swarm Optimization and Genetic Algorithms proved to be efficient, robust and can be considered reliable in obtaining the parameter estimates for Nonlinear Regression Model.

Keywords**:** Gauss Newton, Levenberg – Marquadt, Optimal estimates, Local optimal

1. INTRODUCTION

In statistical modelling, regression analysis is a set of statistical processes for estimating the relationships among variables. It includes many techniques for modeling and analyzing several variables, when the focus is on the relationship between a dependent variable and one or more independent variables. More specifically, regression analysis helps one to understand how the typical value of the dependent variable changes when anyone of the independent variables varies, while the other independent variables are held fixed. Regression analysis is performed so as to determine the correlations between two or more variables having cause-effect relations, and to make predictions for the topic by using the relation. The regression using a single independent variable is called univariate regression analysis while the analysis using more than one independent variable is called multiple or multivariate regression analysis. Genetic algorithm (G.As) using α-level estimation method are promising approach to evaluate the parameter estimation in non-linear regression model. Generally, a regression model can be expressed as:

Y_i= F $(X_i \ \beta)$ + ϵ_i

where Y_i is a response variable, X_i is a vector of explanatory variables, β is a vector of unknown parameters, ϵ_i is the residual error term.

The fuzzy aspect also exists in literature see, [1], [2] and [3].

In recent past, lots of stochastic algorithm have been developed to provide reliable solutions to Nonlinear Regression Models. Numerous researchers, has devoted sufficiently enough article and journals to the study and analysis of parameter estimation of the Nonlinear Regression Model. Genetic Algorithm (GA) are a type of optimization algorithm commonly used to solve optimum experimental problems that are characterized by mixed continuous discrete variables, discontinuous and non-convex design.

The use of standard nonlinear techniques for this type of problems will be inefficient and computationally expensive. These algorithms are well suited for solving such problems, and in most cases they can find the global optimum solution with a high probability without no extra information about the given problem [4]. In GAs, pool or a population of possible solutions to the given problem are produced at a generation. These generations then undergo recombination and mutation

Corresponding Author**:** Onoghojobi B., Email: ngolewe@yahoo.com, Tel: +2348034933133

(like in natural genetics), producing new children, and the process is repeated over various generations. Each individual (single solution) is assigned a fitness value (based on its objective function value) and the "fitter" individuals are given a higher chance to mate and yield more "fitter" individuals. This is in line with Darwinian theory of "Survival of the Fittest".
The basic components common to all forms of genetic algorithms are:

i. a fitness function for optimization
ii. a population of chromosomes
iii. selection of which chromosomes will reproduce
iv. crossover to produce next generation of chromosomes
v. random mutation of chromosomes in new generation

The fitness function is the function that the algorithm is trying to optimize. It is one of the most pivotal parts of the algorithm. The term chromosomes refers to a numerical value or values that represent a candidate's solution to the problem that the genetic algorithm is trying to solve. Als0 [5] emphasized that Nonlinear Regression based prediction has already been successfully implemented in various areas of scientific research and technology. It is mostly used for functional estimation or solutions that enable modeling of a dependence between two variables.

In their study, [6] showed that the Particle swarm optimization and Genetic algorithm produces the same effectiveness. They used Griewangks Function as input test function to compare Particle Swarm Optimization and Genetic algorithm in order to obtain best optimized value. A statistical test was conducted to investigate their efficiency alongside a statistical hypothesis.

Equally, [7] in their works discussed that optimization methods known as meta-heuristics can offer robust methods of finding a solution, and increases the likelihood of converging onto a solution at the global optimum. They stated implicitly that these methods are capable of engaging with numerically noisy optimization problem that can be difficult for gradient based methods.

2. The basic formulation of the Genetic Algorithm

The multiple linear regression model is formulated as follows:

$Y_i \ = \beta_0 \ + \ \beta_p \ X_{ip} \ + \ \epsilon_i \ , i = 1, \dots, n.$

The least squares estimator for the model is

$\hat{\beta} = (X^T X)^{-1} X^T Y$

where

$X = \ (X_1^T \ , \dots, X_n^T)^{\mathrm{T}}$ and $Y = \ (y_1, \dots, y_n).$

The Ordinary Least Squares (OLS) is a technique used to estimate the parameters of the linear regression model. It is noted at this point, the two ways in establishing the OLS regression model:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 \, x_{i1} + \hat{\beta}_2 \, x_{i2} + \dots + \hat{\beta}_p x_{ip} \ = x_i^T \hat{\beta} \ \ , i = 1,2, \dots, n \qquad (1)$$

Firstly,

$\sum_{i=1}^{n} = 1 \, e_i = \sum_{i=1}^{n} (y_i \text{ - } \hat{y}_1) = 0$

where

$e_i \ = y_i \text{ - } \hat{y}_i$ is called the ith residual

Secondly,

$$\hat{e}^T \hat{e} = \ \sum_{i=1}^{n} e_i^2 \ = \ \sum_{i=1}^{n} (y_i - x_i^T \hat{\beta} \ \ )^2 \qquad (2)$$

$= \ \sum_{i=1}^{n} (y_i - x_i^T \hat{\beta} \ \ )^2$;

which can be written in matrix form as;

$(Y - X\hat{\beta})^T (Y - X\hat{\beta}) \ = Y^T \mathrm{Y} - 2Y^T \, X\hat{\beta} + \ \hat{\beta}^T X^T X \hat{\beta}$

Note that $Y^T \, X\hat{\beta} = \ \hat{\beta}^T X^T Y$ since they are both scalars. To find $\hat{\beta} = (\hat{\beta}_0 \, , \hat{\beta}_1 \, , \dots , \hat{\beta}_p)^{\mathrm{T}}$ that minimizes $(\hat{e}^T \hat{e})$, we take the partial derivatives of $(\hat{e}^T \hat{e})$ with respect to $\hat{\beta}$ and equate it to zero to obtain $\hat{\beta} = (\hat{\beta}_0 \, , \hat{\beta}_1 \, , \dots , \hat{\beta}_p)^{\mathrm{T}}$

Here,

$$\frac{\partial \hat{e}^T \hat{e}}{\partial \hat{\beta}} = 0 - 2X^T Y + 2X^T X B \qquad (3)$$

Finally,

$\hat{\beta}_{OLS} = (X^T X)^{-1} X^T Y$

The general form of Nonlinear Regression Models can be written as:

$Y = f(X, \theta) + \varepsilon$

where Y is the dependent variable, X is an (nx1) vector of independent variables, θ is a (kx1) (nonlinear) parameter vector, and ε is a random error.

One of the Nonlinear Regression models widely used in empirical studies is Power Regression model. A regression model of this type containing a single independent variable can be written as follows:

$Y_i = \theta_0 \exp(\theta_1 * X_i) + \varepsilon_i$

where $Y_i$is the dependent variable, $X_i$is the independent variable, $\varepsilon_i$is a stochastic error term, and θ_0 and $\theta_1$are the parameters of the model.

Wald Statistics

The Wald statistics is a parametrical test named after the Transylvanian statistician, Abraham Wald with a great variety of uses. Whenever a relationship within or between data can be expressed as a statistical model with parameters to be estimated from a sample, this test can be used to test the true value of the parameter based on the sample estimate Wald is used to test the significance of each variable and it is simply the Z statistics

$$\text{Wald} = \frac{{\beta_i}^2}{\sigma^2} \qquad (4)$$

If it is a categorical variable, the Wald statistic is computed. The Wald statistic is

$$\text{Wald} = \frac{\beta_2^1}{\sigma^2{}_{\beta_1}} \qquad (5)$$

If it is a categorical variable, the Wald statistics is Chi-square with degree of freedom equal to the number of parameters estimated.

Test for the Odds Ratio in Logistics Regression with Wald Test

In a test, where we can have two binary covariate (X & Z) in the Logistics Regression model, Wald Statistics is used in testing their significance where

Y=Response Variable (First binary covariate)

X=Exposure Variable (Second binary covariate)

Z=Confounder Variable.

The probability of a binary event

$$P_r\left(X = \frac{1}{x\,,z}\right) = \frac{\exp\left(\beta_o + \beta_1 x + \beta_2 z\right)}{1 + \exp\left(\beta_o + \beta_1 x + \beta_2 z\right)} \qquad (6)$$

$$\text{Log}\left[\frac{P_r\left(\frac{y-1}{x,z}\right)}{1 - P_r\left(\frac{y=1}{x,z}\right)}\right] = \beta_o + \beta_1 x + \beta_2 z \qquad (7)$$

where P_0 = baseline probability

$$P_0 = \Pr(y = \frac{1}{x=0,z=0}) = \frac{\exp(\beta o)}{1+\exp(\beta o)} \qquad (8)$$

Power Analysis on Wald Test

The significance of the slope β_1 is commonly tested with the Wald test

$$Z = \frac{\beta_2}{\beta_1} \qquad (9)$$

Therefore, the power for the two sided Wald test given by [8] is

$$\text{Power} = \phi\left(-Z_{1-\frac{\alpha}{2}} + \frac{\beta_1\sqrt{N}}{\sqrt{\nu}}\right) + \phi\left(-Z_{1-\frac{\alpha}{2}} - \frac{\beta_1\sqrt{N}}{\sqrt{\nu}}\right) \qquad (10)$$

where Z is the usual quantile of the standard normal distribution

ν is calculated by taking

P_x as the probability that X = 1 in the sample

P_z as the probability that Z = 1 in the sample regression

Assuming that relationship between X and Z is a logistic regression

$$\Pr(X=1/z) = \frac{\exp\left(\gamma_o+\gamma_1 Z\right)}{1+\exp\left(\gamma_o+\gamma_1 Z\right)} \tag{11}$$

where γ_o value is obtained from

$$\operatorname{Exp}(\gamma_o) = \frac{Q+\sqrt{Q}^2+4p_x \quad x\left(1-p_x)\right)\exp(\gamma_1)}{2(1-p_x)\exp(\gamma_1)} \tag{12}$$

With Q=Px (1+exp(γ_1)) + Pz (1-exp(γ_1))-1. The Information Matrix for this model

$$I=\begin{pmatrix} L+F+J+H & F+H & J+H\ f \\ F+H & F+H & H \\ J+H & H & J+H \end{pmatrix} \tag{13}$$

where

$$L = \frac{\left(1-P_z\right)\exp\left(\beta_o\right)}{\left(1+\exp\left(\gamma_o\right)\right)\left(1+\exp\left(\beta_o+\beta_1\right)\right)^2} \tag{14}$$

$$H = \frac{P_z \exp\left(\beta_o+\beta_1+\beta_2+\gamma_o+8\right)}{\left(1+\exp\left(\gamma_o\right)\right)\left(1+\exp\left(\beta_o+\beta_1\right)\right)^2} \tag{15}$$

$$F = \frac{1-P_Z\exp\left(\beta o+\beta_1+\gamma_0\right)}{(1+\exp(\gamma_0))(1+\exp\left(\beta o+\beta_1\right))^2} \tag{16}$$

$$J = \frac{P_Z\exp\left(\beta o+\beta_2\right)}{(1+\exp(\gamma_0+\gamma_1))(1+\exp\left(\beta o+\beta_1\right))^2} \tag{17}$$

The value of V is the (2,2) element of the inverse of I and the value of the regression co-efficient are input as P_0 and the following odd ratio as follows

Oryx = exp(β_1)

ORyz = exp(β_2)

ORxz = exp(γ_1).

Computation of the Genetic Algorithms

The computational procedure involved in maximizing the fitness function $F(x_1, x_2, x_3, \cdots, x_n)$ in the genetic algorithm can be described by the following steps.

i. Choose a suitable string length $l = nq$ to represent the n design variables of the design vector X. Assume suitable values for the following parameters: population size m, crossover probability pc, mutation probability *pm*, permissible value of standard deviation of fitness values of the population (*sf*) max to use as a convergence criterion, and maximum number of generations (*imax*) to be used an a second convergence criterion.

ii. Generate a random population of size m, each consisting of a string of length $l = nq$. Evaluate the fitness values F_i, $i = 1, 2, \cdots, m$, of the m strings.

iii. Carry out the reproduction process.

iv. Carry out the crossover operation using the crossover probability *pc*.

v. Carry out the mutation operation using the mutation probability *pm* to find the new generation of *m* strings.

vi. Evaluate the fitness values *Fi*, $i = 1, 2, \cdots, m$, of the *m* strings of the new population. Find the standard deviation of the *m* fitness values.

vii. Test for the convergence of the algorithm or process. If (*sf*)*max*, the convergence criterion is satisfied and hence the process may be stopped. Otherwise, go to step viii.

viii. Test for the generation number. If *imax*, the computations have been performed for the maximum permissible number of generations and hence the process may be stopped. Otherwise, set the generation number as $i = i + 1$ and go to step iii.

3. ANALYSIS OF RESULTS

Table 1: The structure of Test-Bed (nonlinear) models

Test Problem	Dataset	Regression Model
1	Chwirut2	y = $\frac{exp[-\beta_1 x]}{\beta_2+\beta_3 x}$
2	DanWood	$y = \beta_1 x^{\beta_2}$
3	Misra1d	$y = \frac{\beta_1\beta_2 x}{1+\beta_2 x}$
4	Eckerle4	$\frac{\beta_1}{\beta_2} exp\left[\frac{-(x-\beta_3)^2}{2\beta_2^2}\right]$
5	Rat42	$y = \frac{\beta_1}{1+exp[\beta_2-\beta_3 x]}$

Table 2: Properties of Test-Bed (Nonlinear model) Problems

S/N	Datasets name	Difficulty Level/Classification	Number of Observation / Parameters
1	Chwirut2	Lower/Exponential	54/3
2	Danwood	Lower/Miscellaneous	6/2
3	Misra1d	Average/Exponential	14/2
4	Eckerle4	Higher/Miscellaneous	35/3
5	Rat42	Higher/Miscellaneous	9/3

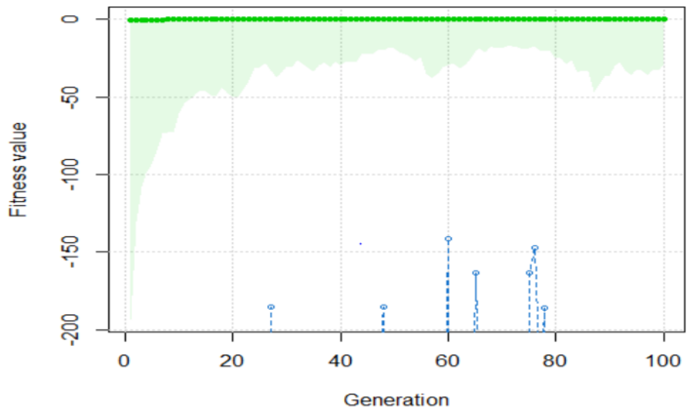

Figure 1: The Residual Sum of Squares behavior of the GA approach for Dan Wood model

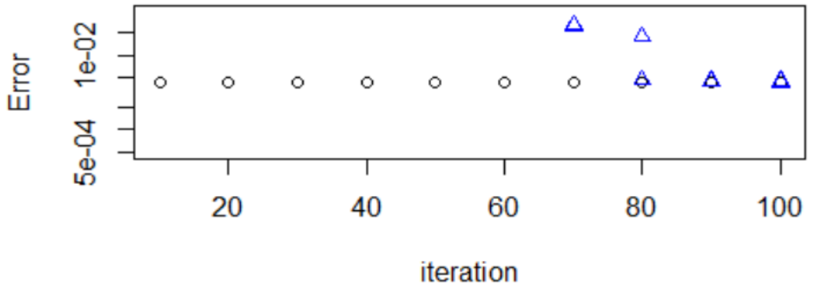

Figure 2: The residual plot.

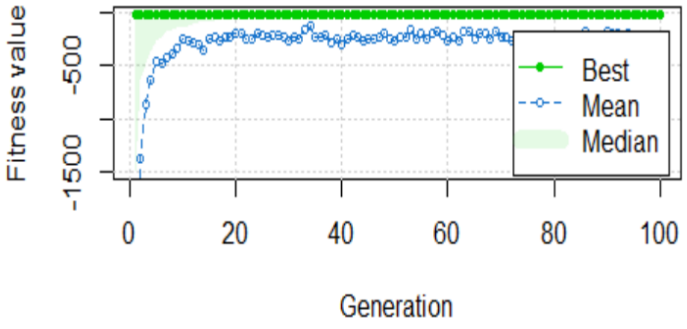

Figure 3: The genetic residual plot.

4. Conclusion

In this paper, the results obtained were displayed on tables and graphs. Particle swarm optimization and Genetic Algorithm proved to be efficient and can be considered reliable in obtaining the parameter estimates for Nonlinear Regression Model.

REFERENCES

[1] Seyedmonir S., Bayrami M. Ghoushchi S. J., Yengnjeh A. A. and Herawi H. M. (2021). Extended fully fuzzy linear regression to analyses a solid cantilever beam moment. Hindawi Mathematical problem in enginnering. Vol. 2021. Article ID 2684816.http://doiorg/10.1155/2021/2684816

[2] Kin H. and Jung H. Y, (2020). Ridge fuzzy regression modelling for solving multicollinearity. MDPI. Mathematic 8, 1572 doi: 10.3390/math809157

[3] Zhang Y. Qu H., Wang W., and Zhao J. (2020). A novel fuzzy time series forecasting model based on multiple linear regression and time series clustering. Hindawi mathematical problems in engineering. Vol. 2020, Article ID 9546792. https://doi.org/10.1155/2020/9546792

[4] Rao, U. (2009). Characteristics, Correlates, and Outcomes of childhood and adolescent depressive disorders. *https://pubdocs. worldbank.org*

[5] Akoa B. E., & Lebowsky, F. (2013). Video decoder monitoring using nonlinear regression. *IEEE 19th International On-line Testing Symposium (IOLTS)*, 175-178.

[6] Chandrashaker Reddy B., Venkat Prasad Reddy P., & Rajeshwari M., Kavya Y. Sai (2017). Correlation of GA and PSO for Analysis of Efficient optimization. *International Journal of Advance Research and Development*. Vol. 2.

[7] Shafi M.A., Rusiman M. S. and Abdullahi S. N. S. (2021) Application of fuzzy linear regression with symmetric parameter for predicting tumor size of colorectal cancer. Mathematics and Statistic (1); 36 -40.

[8] Demidenko, E. (2007). Sample size determination for Logistic Regression. Revisited. *https://www.researchgate.net>6649.*