Journal homepage www.jzs.univsul.edu.iq
Journal of Zankoy Sulaimani
Part-A- (Pure and Applied Sciences)

Parameter Estimation for Binary Logistic Regression Using Different Iterative Methods

Khwazbeen S. Fatah[1]& Rzgar F. Mahmood[2]

1 College of science / Mathematics Department-Salahaddin University, Ierbil-Iraq
Email: khwazbeen.fatah@su.edu.krd; khwazbeen@yahoo.com
2 College of Education / Mathematics Department-Garmian University, Kalar-Sulaimania-Iraq
Email: rzgarfariq@gmail.com; rzgarfariq@yahoo.com

Article info

Original: 12 July 2016
Revised: 7 November 2016
Accepted: 20 November 2016
Published online: 20 June 2017

Key Words:
Binary Logistic Regression,
Maximum Likelihood Estimation,
NRM,
Modified NRM.

Abstract

Logistic Regression Analysis describes how a response variable having two or more categories is associated with a set of predictor variables (continuous or categorical) through a probability function. When the response variable is with only two categories a Binary Logistic Regression Model is the most widely used approach. The main deficiency with this method is in estimating logistic parameters numerically by applying Maximum Likelihood Estimation using Newton Raphson Method. In this paper, in order to improve the efficiency of the parameter estimates, four different modifications D-B-N; C-M-J; A-C-T; and L-W-W-Z, for NRM are introduced; each is an iterative method based on NRM. To specify the efficiency of these approaches, based on the number of iterations, all these procedures are compared with each other and then with NRM to identify the most efficient one. Finally, practical implementations for these procedures are given.

Introduction

Regression analysis is one of the most popular statistical techniques used for analyzing the relationship between a dependent variable and one or more independent variables. Usually, regression analysis is used to reveal and specify the effect of one variable upon another, one of the set variable is called dependent (outcome or response) variable denoted by Y and the other set is called independent (predictor or explanatory) variables denoted by $\boldsymbol{X}$. The linear regression model is, under certain conditions, a valuable tool for quantifying the impact of one or more variables on a response continuous variable. For situations when the response variable is discrete or qualitative other methods have been proposed. One of these is the Logistic Regression Model (LRM) or logit model, which specifically deals with the case of binary (dichotomous) response variable (Pampel 2000; Menard 2002; Cramer 2003; Kleinbaum and Klein 2010; Agresti 2013). Since its development, Binary LRM has become one of the most widely used statistical procedures employed by statisticians and researchers for the analysis of binary and proportional response data in many fields of research studies such as health sciences, business, engineering, the physical sciences, economics, management, life and biological sciences, the social sciences, and many other disciplines (Montgomery and Peck 1982; Yarandiand and Simpson 1991; Reilly et al. 2009; Hassanien et al. 2014).

For estimating logistic parameters, Maximum Likelihood Estimation (MLE) is normally used (Hosmer and Lemeshow 2013). However, in applying MLE a system of non-linear equations is obtained; the solution of such system is not easily derived algebraically (Albert and Anderson 1984). In order to solve such system to get numerically estimated solutions, iterative methods are used. One of the most popular approximated method is Newton Raphson Method (NRM), which is a well-known iterative formula used to find the roots of nonlinear equations (Ben-Israel 1966; Green 1984; Kelley 1995; Serfling 1980; Mak 1993; Givens and Hoeting 2013; Demira and Akkusb 2015).

In this paper, following introduction, binary logistic regression and parameter estimation methods are introduced. Next, NRM and all four modifications for NRM are introduced along with their procedures and then these algorithms are compared. Finally, the conclusion for this study is given.

The Binary Logistic Regression and Parameters Estimation Method

Suppose a binary random variable $\boldsymbol{y}$ where $= (y_1, y_2, \dots, y_N)^T$, each y_i, $i = 1, \dots, N$ follows a Bernoulli distribution that y_i take either the value 1 or the value 0 with probability $\pi_i(x), 1 - \pi_i(x)$ respectively where $\pi_i(x)$ is the mean of the binary variable representing the conditional probability $p(y_i = 1|\boldsymbol{x})$ and $\boldsymbol{x} = (x_{i0}, x_{i1}, x_{i2}, \dots, x_{ik}) \in \mathbb{R}^{K+1}$ is a vector of $K + 1$ explanatory variables.

The LRM relates $\pi_i(x)$ through a logistic function, to variation in the explanatory variables.

The described from of the LRM with unknown parameters $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_K)^T$ with $x_{i0} = 1$ is given by

$$\pi_i(x) = \frac{e^{\sum_{k=0}^{K} \beta_k x_{ik}}}{1 + e^{\sum_{k=0}^{K} \beta_k x_{ik}}} \tag{1}$$

the specific transformation of $\pi_i(x)$ is called the Logit transformation; it is given by

$$\text{logit}\big(\pi_i(x)\big) = \ln\left(\frac{\pi_i(x)}{1-\pi_i(x)}\right)$$

Equation (1) can be rewritten as

$$\text{logit}\big(\pi_i(x)\big) = \boldsymbol{X}^T\boldsymbol{\beta}$$

If a sample of N −independent observation $\{(y_i, x_i)\}_{i=1,\dots,N} \in [\{0,1\} \times \mathbb{R}^{K+1}]^N$ where y_i represents a binary outcome value, and x_i is the predictor value for the i^{th} subject (Rashid, 2008).

To find the MLE for $\boldsymbol{\beta}$, the likelihood function $L(\boldsymbol{\beta})$ is defined as

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{N} \pi_i(x)^{y_i}\big(1 - \pi_i(x)\big)^{n_i - y_i} \tag{2}$$

Substituting (1) in (2), obtain:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{N} \left(\frac{e^{\sum_{k=0}^{K} \beta_k x_{ik}}}{1 + e^{\sum_{k=0}^{K} \beta_k x_{ik}}}\right)^{y_i} \left(1 - \frac{e^{\sum_{k=0}^{K} \beta_k x_{ik}}}{1 + e^{\sum_{k=0}^{K} \beta_k x_{ik}}}\right)^{n_i - y_i}$$

$$= \prod_{i=1}^{N} \left(e^{y_i \sum_{k=0}^{K} \beta_k x_{ik}}\right)\left(1 + e^{\sum_{k=0}^{K} \beta_k x_{ik}}\right)^{n_i} \tag{3}$$

Taking the natural log of (3) yields the log likelihood function:

$$l(\boldsymbol{\beta}) = \log\big(L(\boldsymbol{\beta})\big) = \sum_{i=1}^{N} \left[y_i \left(\sum_{k=0}^{K} \beta_k x_{ik} \right) - n_i \log\left(1 + e^{\sum_{k=0}^{K} \beta_k x_{ik}}\right) \right] \tag{4}$$

To find the critical points of $l(\boldsymbol{\beta})$

$$\frac{\partial l(\boldsymbol{\beta})}{\partial \beta_k} = \sum_{i=1}^{N} [y_i x_{ik} - n_i \pi_i(x) x_{ik}] = 0 \tag{5a}$$

Then, equation (\ref{eq:ch3:3.5a}) can be expressed as a matrix multiplication as follows:

$$G(\boldsymbol{\beta}) = l'(\boldsymbol{\beta}) = \boldsymbol{X}^T\boldsymbol{\beta} \tag{5b}$$

Where $l'(\boldsymbol{\beta})$ is a column vector of length $K+1$ whose elements are $\frac{\partial l(\boldsymbol{\beta})}{\partial \beta_k}, k = 0,1,\dots,K$, likewise $\boldsymbol{\mu}$ is a column vector of length N with elements $\mu_i = n_i\pi_i$, i.e. $\boldsymbol{\mu} = (n_1\pi_1, n_2\pi_2, \dots, n_N\pi_N)^T$, and $\boldsymbol{X}^T$ is a $(K+1)\times N$ matrix.

The ML estimates for $\boldsymbol{\beta}$ can be found by setting each of the $K+1$ equations in (5) equal to zero and solving for each β_k. In order to maximize the estimates each of $K+1$ equations are differentiated with respect to β_k, denoted by $\beta_{k'}$. The general form of the matrix of second partial derivatives is

$$\frac{\partial^2 l(\boldsymbol{\beta})}{\partial\beta_k\partial\beta_{k'}} = \frac{\partial l(\boldsymbol{\beta})}{\partial\beta_{k'}}\sum_{i=1}^{N}[y_i x_{ik} - n_i\pi_i(x)x_{ik}]$$
$$= -\sum_{i=1}^{N} n_i x_{ik}\pi_i(x)\big(1-\pi_i(x)\big)x_{ik'} \tag{6a}$$

Where $k = 0,1,\dots,K$ and $k' = 0,1,\dots.,K$.

Equation (6a) can be expressed in term of matrix multiplication $H(\boldsymbol{\beta})$

$$H(\boldsymbol{\beta}) = -\boldsymbol{X}^T D\boldsymbol{X}$$

$H(\boldsymbol{\beta})$ is called Hessian matrix, where D is $N\times N$ diagonal matrix.

Setting the equations in (5) equal to zero results in a system of $K+1$ nonlinear equations each with $K+1$ unknown variables. The solution of the system is a vector with elements, β_k.

After verifying that the matrix of second partial derivatives is negative definite, and that the solution is the global maximum rather than a local maximum, then we can conclude that this vector contains the parameter estimates for which the observed data would have the highest probability of occurrence (Van Den Berg, et. al. 1984).

However, solving this system of nonlinear equations, described by (5), is not easy; the solution cannot be derived algebraically as it can be in the case of linear equations. For solving such nonlinear system, the classical NRM has been applied to reach at the approximated estimates for parameters.

In this study, in order to improve parameter estimates resulting from applying NRM, various iterative techniques are studied. These methods are modifications for NRM but with more efficient results.

In the following, NRM is first introduced and then the four iterative methods are proposed; they are modification for NRM and applied to the system of equations described by (5).

Newton-Raphson Method

Consider the system of nonlinear equations $G(\boldsymbol{\beta}) = 0$, for solving this system, if an initial guess $\boldsymbol{\beta}^{(0)}$ is computed from the observed data, then

$$\boldsymbol{\beta}^{(r+1)} = \boldsymbol{\beta}^{(r)} - \left[H\big(\boldsymbol{\beta}^{(r)}\big)\right]^{-1} G\big(\boldsymbol{\beta}^{(r)}\big) \tag{7}$$

Here $G'(\boldsymbol{\beta}^{(r)})$ refers to the derivative $G(\boldsymbol{\beta}^{(r)})$.

Equation (7) is called the NR formula for solving system of nonlinear equations of the form $G(\boldsymbol{\beta}^{(r)}) = 0$. For solving such system using (7), an initial guess for the root is computed. Assume that $\boldsymbol{\beta}^{(0)}$ is that guess, the next value $\boldsymbol{\beta}^{(1)}$ will be determined. This iterative process will continue until the desirable root is obtained; it is with $||\boldsymbol{\beta}^{(r+1)} - \boldsymbol{\beta}^{(r)}|| < \varepsilon$ for some specific value $\varepsilon, r = 0,1,\dots..$

Remark:

i. $G'(\boldsymbol{\beta}) = H(\boldsymbol{\beta})$.

ii. An initial guess $\boldsymbol{\beta}^{(0)} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Q}$, is determinate from (5), $\boldsymbol{Q}$ is a column vector with length N; it is calculated from the observed data and has elements $Q_i = \log\left(\frac{y_i}{n_i - y_i}\right)$.

The following algorithm shows the procedure for NRM applied to (5).

Algorithm 1

For solving the system $G(\boldsymbol{\beta}^{(r)}) = 0$, the main steps for the algorithm is summarized as:

i. Set $r = 0$: Find an initial guess $\boldsymbol{\beta}^0$;

ii. Compute $H'(\boldsymbol{\beta}^r)$;

iii. Evaluate

$$\boldsymbol{\beta}^{(r+1)} = \boldsymbol{\beta}^{(r)} - \left[H\left(\boldsymbol{\beta}^{(r)}\right)\right]^{-1} G\left(\boldsymbol{\beta}^{(r)}\right);$$

iv. If $||\boldsymbol{\beta}^{(r+1)} - \boldsymbol{\beta}^{(r)}|| > \varepsilon$, then $r = r + 1$ and return to ii;

v. Stop.

Modified Newton Methods

The main goal in this work is to improve the efficiency of the parameter estimates when solving the system $G(\boldsymbol{\beta}^{(r)}) = 0$ which is obtained by NRM. For this purpose, four different iterative methods are proposed; each one is a new modification for NRM, but can approach at the true value with less iterations than that in case of NRM.

The four iterative methods proposed in this study are introduced below:

D-B-N Method

For this method, to solve the nonlinear system $G(\boldsymbol{\beta}^{(r)}) = 0$, an iterative technique proposed by Darvishi and Barati (Darvishi and Barati (2007)) is applied, which is a modification for NRM.

In this method, first by applying NRM, a root $\boldsymbol{\alpha}^{(0)}$ at an initial guess $\boldsymbol{\beta}^{(0)}$ is found. Then, to compute a new root, say $\boldsymbol{\beta}^{(1)}$, D-B-N method is implemented. The process continues until an efficient estimate is obtained.

The general form D-B-N method, for solving the system $G(\boldsymbol{\beta}^{(r)}) = 0$, is

$$\boldsymbol{\beta}^{(r+1)} = \boldsymbol{\beta}^{(r)} - \left[H\left(\boldsymbol{\beta}^{(r)}\right)\right]^{-1} \left(G\left(\boldsymbol{\beta}^{(r)}\right) + G\left(\boldsymbol{\alpha}^{(r)}\right)\right) \tag{8}$$

where

$$\boldsymbol{\alpha}^{(r)} = \boldsymbol{\beta}^{(r)} - \left[H\left(\boldsymbol{\beta}^{(r)}\right)\right]^{-1} G\left(\boldsymbol{\beta}^{(r)}\right) \tag{9}$$

Therefore, each $\boldsymbol{\alpha}^{(r)}$ is computed by applying NRM at $\boldsymbol{\beta}^{(r)}$ with an initial guess $\boldsymbol{\beta}^{(0)}$ computed first, $= 0,1, \ldots .$

The main steps for this method is described by the following algorithm.

Algorithm 2

The following step explains the algorithm for solving the nonlinear system $G(\boldsymbol{\beta}^{(r)}) = 0$:

i. Set $r = 0$: Find an initial guess $\boldsymbol{\beta}^0$;

ii. Compute $H'(\boldsymbol{\beta}^r)$;

iii. Evaluate

$$\boldsymbol{\alpha}^{(r)} = \boldsymbol{\beta}^{(r)} - \left[H\left(\boldsymbol{\beta}^{(r)}\right)\right]^{-1} G\left(\boldsymbol{\beta}^{(r)}\right);$$

iv. Determine parameter estimates

$$\boldsymbol{\beta}^{(r+1)} = \boldsymbol{\beta}^{(r)} - \left[H\left(\boldsymbol{\beta}^{(r)}\right)\right]^{-1} \left(G\left(\boldsymbol{\beta}^{(r)}\right) + G\left(\boldsymbol{\alpha}^{(r)}\right)\right);$$

v. If $||\boldsymbol{\beta}^{(r+1)} - \boldsymbol{\beta}^{(r)}|| > \varepsilon$, then $r = r + 1$ and return to ii;

vi. Stop.

C-M-T Method

The second iterative procedure for solving the system $G(\boldsymbol{\beta}^{(r)}) = 0$, is a technique proposed by Cordero-Torregrosa (Cordero A., et al., (2009)).

In this method as for D-B-N technique, the solution for the system is performed by applying NRM first to obtain $\boldsymbol{\alpha}^{(r)}$ and then using $\boldsymbol{\alpha}^{(r)}$ to compute $\boldsymbol{\gamma}^{(r)}$, where

$$\boldsymbol{\alpha}^{(r)} = \boldsymbol{\beta}^{(r)} - \left[H\left(\boldsymbol{\beta}^{(r)}\right)\right]^{-1} G\left(\boldsymbol{\beta}^{(r)}\right) \tag{10}$$

and

$$\boldsymbol{\gamma}^{(r)} = \boldsymbol{\alpha}^{(r)} - \left[-2H\left(\boldsymbol{\beta}^{(r)}\right) + 6H\left(\boldsymbol{\alpha}^{(r)}\right)\right]^{-1}\left[H\left(\boldsymbol{\beta}^{(r)}\right) + 3H\left(\boldsymbol{\alpha}^{(r)}\right)\right]\left\{H\left(\boldsymbol{\beta}^{(r)}\right)\right]^{-1} G\left(\boldsymbol{\beta}^{(r)}\right)$$

Hence, $\boldsymbol{\alpha}^{(r)}$ and $\boldsymbol{\gamma}^{(r)}$ are used to compute approximated values for the root $\boldsymbol{\beta}^{(r+1)}$, where

$$\boldsymbol{\beta}^{(r+1)} = \boldsymbol{\beta}^{(r)} - \left[-H\left(\boldsymbol{\beta}^{(r)}\right) + 3H\left(\boldsymbol{\alpha}^{(r)}\right)\right]^{-1} G\left(\boldsymbol{\gamma}^{(r)}\right) \tag{11}$$

$\boldsymbol{\beta}^{(0)}$ is an initial guess for the process; $r = 0,1, \dots$.

The algorithm describing this approach is given below.

Algorithm 3

Algorithm 3 display the main steps for C-M-T method:

i. Set $r = 0$: Find an initial guess $\boldsymbol{\beta}^{0}$;

ii. Compute $H'(\boldsymbol{\beta}^{r})$;

iii. Evaluate

$$\boldsymbol{\alpha}^{(r)} = \boldsymbol{\beta}^{(r)} - \left[H\left(\boldsymbol{\beta}^{(r)}\right)\right]^{-1} G\left(\boldsymbol{\beta}^{(r)}\right);$$

$$\boldsymbol{\gamma}^{(r)} = \boldsymbol{\alpha}^{(r)} — \left[2H\left(\boldsymbol{\beta}^{(r)}\right) + 6H\left(\boldsymbol{\beta}^{(r)}\right)\right]^{-1}\left[H\left(\boldsymbol{\beta}^{(r)}\right) + 3H\left(\boldsymbol{\beta}^{(r)}\right)\right]\left[H\left(\boldsymbol{\beta}^{(r)}\right)\right]^{-1} G(\boldsymbol{\beta}^{(r)});$$

iv. Determine parameter estimates

$$\boldsymbol{\beta}^{(r+1)} = \boldsymbol{\beta}^{r} - \left[-H(\boldsymbol{\beta}^{r}) + 3H\left(\boldsymbol{\alpha}^{(r)}\right)\right]^{-1} G\left(\boldsymbol{\gamma}^{(r)}\right);$$

v. If $||\boldsymbol{\beta}^{(r+1)} - \boldsymbol{\beta}^{(r)}|| > \varepsilon$, then $r = r + 1$ and return to ii;

vi. Stop.

A-C-T Method

This approach is another iterative technique proposed by Abad-Torregrosa (Abad, et al. (2013)); it is based on the D-B-N method.

In this method, the iterative formula, for solving $G(\boldsymbol{\beta}) = 0$, is conducted when $\boldsymbol{\alpha}^{(r)}$ is a solution obtained by Newton's method and $\boldsymbol{\gamma}^{(r)}$ is another solution which is found by applying D-B-N method, for initial guess $\boldsymbol{\beta}^{(0)},\ r = 0,1, \dots$.

Therefore, the procedure for this method is as follows:
First

$$\boldsymbol{\alpha}^{(r)} = \boldsymbol{\beta}^{(r)} - \left[H\left(\boldsymbol{\beta}^{(r)}\right)\right]^{-1} G\left(\boldsymbol{\beta}^{(r)}\right) \tag{12}$$

and then

$$\boldsymbol{\gamma}^{(r)} = \boldsymbol{\beta}^{(r)} — \left[H\left(\boldsymbol{\beta}^{(r)}\right)\right]^{-1} \left(G\left(\boldsymbol{\beta}^{(r)}\right) + G\left(\boldsymbol{\alpha}^{(r)}\right)\right) \tag{13}$$

Finally, the next parameter estimate found by A-C-T method, at a point $\boldsymbol{\alpha}^{(r)}$ and $\boldsymbol{\gamma}^{(r)}$, as

$$\boldsymbol{\beta}^{(r+1)} = \boldsymbol{\gamma}^{(r)} — \left[H\left(\boldsymbol{\alpha}^{(r)}\right)\right]^{-1} G\left(\boldsymbol{\gamma}^{(r)}\right) \tag{14}$$

The main steps for the procedure is summarized by the following algorithm:

Algorithm 4

The following steps, display how to find parameter estimates by A-C-T method:

i. Set $r = 0$: Find an initial guess $\boldsymbol{\beta}^0$;
ii. Compute $H'(\boldsymbol{\beta}^r)$;
iii. Evaluate

$$\boldsymbol{\alpha}^{(r)} = \boldsymbol{\beta}^{(r)} - \left[H\left(\boldsymbol{\beta}^{(r)}\right)\right]^{-1} G\left(\boldsymbol{\beta}^{(r)}\right);$$

$$\boldsymbol{\gamma}^{(r)} = \boldsymbol{\beta}^{(r)} — \left[H\left(\boldsymbol{\beta}^{(r)}\right)\right]^{-1} \left(G\left(\boldsymbol{\beta}^{(r)}\right) + G\left(\boldsymbol{\alpha}^{(r)}\right)\right);$$

iv. Determine parameter estimates

$$\boldsymbol{\beta}^{(r+1)} = \boldsymbol{\gamma}^{(r)} — \left[H\left(\boldsymbol{\alpha}^{(r)}\right)\right]^{-1} G\left(\boldsymbol{\gamma}^{(r)}\right);$$

v. If $||\boldsymbol{\beta}^{(r+1)} - \boldsymbol{\beta}^{(r)}|| > \varepsilon$, then $r = r + 1$ and return to ii;
vi. Stop.

L-W-W-Z Method

The fourth modification for NRM is introduced; it is proposed by Li-Zhang (Li X., et al., (2013)). In this method, the same technique of D-B-N is applied, and then the iterative steps of L-W-W-Z is applied to obtain new estimates.

In this method, if $\boldsymbol{\alpha}^{(r)}$ is obtained from NRM, using an initial guess $\boldsymbol{\beta}^{(r)}$ as

$$\boldsymbol{\alpha}^{(r)} = \boldsymbol{\beta}^{(r)} - \left[H\left(\boldsymbol{\beta}^{(r)}\right)\right]^{-1} G\left(\boldsymbol{\beta}^{(r)}\right) \quad (15)$$

then $\boldsymbol{\alpha}^{(r)}$ is used to compute a value for $\boldsymbol{\gamma}^{(r)}$ from the following

$$\boldsymbol{\gamma}^{(r)} = \boldsymbol{\beta}^{(r)} — \left[H\left(\boldsymbol{\beta}^{(r)}\right)\right]^{-1} \left(G\left(\boldsymbol{\beta}^{(r)}\right) + G\left(\boldsymbol{\alpha}^{(r)}\right)\right) \quad (16)$$

Hence, a new guess for $\boldsymbol{\beta}^{(r+1)}, r = 0,1, \ldots$ can be determent from (16) where

$$^{1)} = \boldsymbol{\gamma}^{(r)} — \left[H\left(\boldsymbol{\gamma}^{(r)}\right)\right]^{-1} G\left(\boldsymbol{\gamma}^{(r)}\right) - \left[H\left(\boldsymbol{\gamma}^{(r)}\right)\right]^{-1} G\left(\boldsymbol{\gamma}^{(r)} - \left[H\left(\boldsymbol{\gamma}^{(r)}\right)\right]^{-1} G\left(\boldsymbol{\gamma}^{(r)}\right)\right) \quad (17)$$

The algorithm that can describe these iterative steps is shown below:

Algorithm 5

The following steps, display how to find parameters estimate by A-C-T method:

i. Set $r = 0$: Find an initial guess $\boldsymbol{\beta}^0$;
ii. Compute $H'(\boldsymbol{\beta}^r)$;
iii. Evaluate

$$\boldsymbol{\alpha}^{(r)} = \boldsymbol{\beta}^{(r)} - \left[H\left(\boldsymbol{\beta}^{(r)}\right)\right]^{-1} G\left(\boldsymbol{\beta}^{(r)}\right);$$

$$\boldsymbol{\gamma}^{(r)} = \boldsymbol{\beta}^{(r)} — \left[H\left(\boldsymbol{\beta}^{(r)}\right)\right]^{-1} \left(G\left(\boldsymbol{\beta}^{(r)}\right) + G\left(\boldsymbol{\alpha}^{(r)}\right)\right);$$

iv. Determine parameter estimates

$$\boldsymbol{\beta}^{(r+1)} = \boldsymbol{\gamma}^{(r)} — \left[H\left(\boldsymbol{\gamma}^{(r)}\right)\right]^{-1} G\left(\boldsymbol{\gamma}^{(r)}\right) - \left[H\left(\boldsymbol{\gamma}^{(r)}\right)\right]^{-1} G\left(\boldsymbol{\gamma}^{(r)} - \left[H\left(\boldsymbol{\gamma}^{(r)}\right)\right]^{-1} G\left(\boldsymbol{\gamma}^{(r)}\right)\right);$$

v. If $||\boldsymbol{\beta}^{(r+1)} - \boldsymbol{\beta}^{(r)}|| > \varepsilon$, then $r = r + 1$ and return to ii;
vi. Stop.

Numerical example

To compare the four iterative numerical methods namely D-B-N, C-M-J, A-C-T, and L-W-W-Z studied in this paper compared to the classical NRM, more than one example have been implemented; one of such examples considered here is Example (1). In this example, the stopping criterion is when $\varepsilon = 10^{-6};\ 10^{-10}$, the iterative proceses terminate when,

i. $||\boldsymbol{\beta}^{(r+1)} - \boldsymbol{\beta}^{(r)}|| < \varepsilon$
ii. $||G(\boldsymbol{\beta}^{(r)})|| < \varepsilon$

Matlab program has been used to perform all algorithms; Algorithm 1-5.

Application

The numerical example in Table (1), which is given by Agresti (2013), is used for the implementation of all algorithm. This data was reported by Cornfield (1962) for a sample of male residents of Framingham, Massachusetts, aged 40-59, classified into 8 subgroups according to blood pressure. During a six-year follow-up period, they were classified according to whether they developed coronary heart disease. This is the response variable. The explanatory variable in the model is the value, x_i which represents the blood pressure in subgroup $i, i = 1,2,...,8$.

Table-1: Cross-Classification of Framingham Men by Blood Pressure and Heart Disease.

			Heart Disease		
$\boldsymbol{i}$	Blood Pressure	x_i	Present y_i	Absent $(n_i - y_i)$	n_i
1	< 117	111.5	3	153	156
2	$117 - 126$	121.5	17	135	252
3	$127 - 136$	131.5	12	272	284
4	$137 - 146$	141.5	16	255	271
5	$147 - 156$	151.5	12	127	139
6	$157 - 166$	161.5	8	77	85
7	$167 - 186$	176.5	16	83	99
8	> 186	191.5	12	35	47

Solution: Assume that the LRM fitted for data in table (1) is

$$Log\left(\frac{\pi_i(x)}{1-\pi_i(x)}\right) = \beta_0 + \beta_1 x_i$$

To estimate parameters β_0 and β_1 by applying the four iterative methods in addition to NRM. If $||\boldsymbol{\beta}^{(r+1)} - \boldsymbol{\beta}^{(r)}|| < \varepsilon$, then the numerical solution are displayed in the following tables (Table (2) and Table (4)).

i. If $\varepsilon = 10^{-6}$, then

Table-2: Numerical Result for Example 1 where

$$\boldsymbol{\beta}_0^{(0)} = \left(\beta_0^{(0)}, \beta_1^{(0)}\right) = (-6.489659818528843, 0.026725954325005\ )^T$$

Numerical Methods	**Iterative**	**Approximation Solution**
NM	204	$(-6.0820534111728515, 0.0243383610352202)^T$
D-B-N	110	$(-6.0820425954692698, 0.0243382980036358)^T$
C-M-J	90	$(-6.0820408090510698, 0.0243382875927755)^T$
A-C-T	76	$(-6.0820396420328091, 0.0243382807916445)^T$
L-W-W-Z	59	$(-6.0820376433997430, 0.0243382691440412)^T$

The following table represents the same table (2) where each value is approximated to 6 decimal place.

Table-3: Numerical Result for Example 1 where

$$\boldsymbol{\beta}_0^{(0)} = \left(\beta_0^{(0)}, \beta_1^{(0)}\right) = (-6.489659818528843, 0.026725954325005\ )^T$$

Numerical Methods	**Iterative**	**Approximation Solution**
NM	204	$(-6.082053, 0.024338)^T$
D-B-N	110	$(-6.082042, 0.024338)^T$
C-M-J	90	$(-6.082040, 0.024338)^T$
A-C-T	76	$(-6.082039, 0.024338)^T$
L-W-W-Z	59	$(-6.082037, 0.024338)^T$

Therefore,

$$Log\left(\frac{\pi_i(x)}{1-\pi_i(x)}\right) = -6.0820 + 0.0243 x_i$$

ii. If $\varepsilon = 10^{-6}$, then

Table-4: Numerical Result for Example 1 where
$\boldsymbol{\beta}_0^{(0)} = \left(\beta_0^{(0)}, \beta_1^{(0)}\right) = (-6.489659818528843, 0.026725954325005\ )^T$

Numerical Methods	Iterative	Approximation Solution
NM	204	$(-6.082033, 0.024338)^T$
D-B-N	110	$(-6.082033, 0.024338)^T$
C-M-J	90	$(-6.082033, 0.024338)^T$
A-C-T	76	$(-6.082033, 0.024338)^T$
L-W-W-Z	59	$(-6.082033, 0.024338)^T$

Therefore,

$$Log\left(\frac{\pi_i(x)}{1-\pi_i(x)}\right) = -6.0820 + 0.0243x_i$$

Results and Discussion

The results given by both tables, (Table (3) and Table (4)) shows that L-W-W-Z method is more efficient than the other three techniques, including NRM when applied to the system of equations described by (5); this method computes an approximated value for the parameter estimate with less number of iterations for both cases when $\varepsilon = 10^{-6};\ 10^{-10}$.

Remark:

The implementation of these iterative methods to the same system using another set of data given by Hosmer (2013) Example 1.1 (page 2) resulted out the same conclusions.

Conclusion

Binary LRM is a widely used approach for analyzing categorical data, the main deficiency with this method is in estimating parameters numerically. Normally, to estimate logistic parameters, MLE is applied. Despite its simplicity, using MLE ends up with a system of non-linear equation, which is known to be a complicated system. In order to solve such non-linear system of equation, numerical methods are used. One of the most popular numerical method, which is an iterative method, is NRM.

In this paper, four different modifications, D-B-N; C-M-J; A-C-T; and L-W-W-Z, for NRM are proposed and their algorithms are introduced. All these approaches are iterative methods too; they are modifications for NRM but each with a new technique. Then, to specify the most efficient method including NRM, all these new procedures are compared with each other and then with NRM using a numerical example as an application. From the implementation of the given example it is shown that the fourth method, namely L-W-W-Z method, is the one that approaches the true value with less iteration. The same procedure for these methods has been applied to another data system and the same conclusion has been obtained.

In this paper, only Binary LRM is considered; it is recommended here to extend these proposed procedures to the case when multinomial regression model is included.

References

[1] Abad, M.; Cordero, A. and Torregrosa, J., "*Fourth- and Fifth-Order Methods for Solving Nonlinear Systems of Equations: An Application to the Global Positioning System*", Abstract and Applied Analysis, Article ID 586708. (2013).

[2] Agresti, A., *"Categorical Data Analysis"*, 3rd edition. John Wiley & Sons, (2013).

[3] Albert, A. and Anderson, J.,"*On the Existence of Maximum Likelihood Estimates in Logistic Regression Models*", Biometrika, Vol. 71, No. 1, pp. 1-10, (1984).

[4] Ben-Israel, A., "*A Newton-Raphson Method for the Solution of Systems of Equations*", Journal of Mathematical Analysis and Applications Vol. 15, No. 2, pp. 243-252, (1966).

[5] Cordero, A.; Martínez, E. and Torregrosa, J., "*Iterative Methods of Order Four and Five for Systems of Nonlinear Equations*", Journal of Computation and Applied Mathematic Vol. 231, No. 2, pp. 541-551, (2009).

[6] Cramer J., *"Logit Models from Economics and Other Fields*", Cambridge University Press, (2003).

[7] Czepiel, S., *"Maximum Likelihood Estimation of Logistic Regression Models: Theory and Implementation*", Available at czep.net/stat/mlelr.pdf. (2002).

[8] Darvishi, M. and Barati, A., *"A third-order Newton-type method to solve systems of nonlinear equations*", Applied Mathematics and Computation Vol. 187, No. 2, pp. 630-635, (2007).

[9] Demira, E. and Akkus, Ö, "*An Introductory Study on "How the Genetic Algorithm Works in the Parameter Estimation of Binary Logit Model?*", International Journal of Sciences: Basic and Applied Research (IJSBAR) Vol. 19, No, 2, pp. 162-180, (2015).

[10] Givens, G. and Hoeting, J., "*Computational Statistics*", John Wiley & Sons, (2013).

[11] Green, P., *"Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives*", Journal of the Royal Statistical Society. Series B (Methodological) Vol. 46, pp. 149-192, (1984).

[12] Hassanien, A.E., Tolba, M. and Azar, A.T. eds., *"Advanced Machine Learning Technologies and Applications: Second International Conference, AMLTA 2014, Cairo, Egypt, November 28-30, 2014. Proceedings*", Vol. 488, Springer, (2014).

[13] Hosmer, D.; Lemeshow. S. and Sturdivant. R., "*Applied Logistic Regression",* 3rd edition. John Wiley & Sons. (2013).

[14] Kelley, C., "*Iterative Methods for Linear and Nonlinear Equations*", Society for Industrial and Applied Mathematics, (1995).

[15] Kleinbaum, D. and Klein, M., "*Logistic Regression: A Self-Learning Text",* 3rd Edition. Springer Science & Business Media, (2010).

[16] Li, X.; Wu, Z.; Wang, L. and Zhang, Q., "*A Ninth-Order Newton-Type Method to solve system of nonlinear equations*". International Journal of Research and Reviews in Applied Science Vol. 16, No. 2, pp. 224-228, (2013).

[17] Mak, T., "*Solving Non-Linear Estimation Equations*", Journal of the Royal Statistical Society. Series B (Methodological) Vol. 55, No. 4, pp. 945-955, (1993).

[18] Menard, S., "*Applied Logistic Regression Analysis*", 2nd edition. Sage, (2002).

[19] Montgomery D. and Peck E., "*Introduction to Linear Regression Analysis",* John Wiley & Sons. Inc, (1982).

[20] Pampel, F., *"Logistic Regression: A Primer*", Sage, (2000).

[21] Rashid, M., "*Inference on Logistic Regression Models*", (Doctoral dissertation, Bowling Green State University), (2008).

[22] Reilly, S., Onslow, M., Packman, A., Wake, M., Bavin, E., Prior, M., Eadie, P., Cini, E., Bolzonello, C. and Ukoumunne, O., "*Predicting stuttering onset by the age of 3 years: A prospective, community cohort study",* Pediatrics, Vol. 123, No. 1, pp. 270–277, (2009).

[23] Serfling, R., *"Approximation Theorems of Mathematical Statistics*", Wiley Series in Probability and Statistics, (1980).

[24] Van Den Berg, C.; Christensen, J. and Ressel, P., *"Harmonic Analysis on Semigroups: Theory of Positive Definite and Related Functions"*, Springer Science & Business Media, (1984).

[25] Yarandi, H. and Simpson, S., *"The Logistic Regression Model and the Odds of Testing HIV positive*", Nursing Research, Vol. 40, No. 6, pp. 372-373, (1991).