



Properties of the Maximum Likelihood Estimates and Bias Reduction for Logistic Regression Model

Nuri H. Salem Badi

Mathematics Department, Faculty of Art and Science—Alabyar, University of Benghazi, Benghazi, Libya
Email: nuri.badi@uob.edu.ly

How to cite this paper: Badi, N.H.S. (2017) Properties of the Maximum Likelihood Estimates and Bias Reduction for Logistic Regression Model. *Open Access Library Journal*, 4: e3625.
<https://doi.org/10.4236/oalib.1103625>

Received: April 21, 2017

Accepted: May 21, 2017

Published: May 24, 2017

Copyright © 2017 by author and Open Access Library Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

A frequent problem in estimating logistic regression models is a failure of the likelihood maximization algorithm to converge. Although popular and extremely well established in bias correction for maximum likelihood estimates of the parameters for logistic regression, the behaviour and properties of the maximum likelihood method are less investigated. The main aim of this paper is to examine the behaviour and properties of the parameters estimates methods with reduction technique. We will focus on a method used a modified score function to reduce the bias of the maximum likelihood estimates. We also present new and interesting examples by simulation data with different cases of sample size and percentage of the probability of outcome variable.

Subject Areas

Applied Statistical Mathematics

Keywords

Logistic Regression Model, Maximum Likelihood Method, Convergence Problems, Bias Reduction Technique

1. Introduction

The logistic regression methods are often used to interpret the statistical analysis of dichotomous outcome variables. It is commonly applied procedure for describing the relationship between a binary outcome variable. The general method of estimating the logistic regression parameter is maximum likelihood (ML). In a very general sense the ML method yields values for the unknown parameters that maximize the probability of the observed set of data. The commonly problem with using ML method is convergence problem, which

occurs when the maximum likelihood estimates (MLE) do not exist. The subject of the assessment behaviour of MLE for logistic regression model is important, as the logistic model is widely used in medical statistics. Much work discusses on logistic regression model address converges problem like [1] or the bias reduction like [2] [3]. Many assumptions and more details considered about the distribution of the coefficients estimated by MLE approach and bias reduction technique, and also for more application and effects of the sample size, see [4] [5]. However, the behavior and properties of bias correction methods are less investigated. A recent paper takes the bias correction technique proposed by [2] to achieve the MLE existing. In the present paper, it centers to evaluate the behavior and properties of the bias reduction method by simulated data with different sample sizes and parameters. The next section, explains the shape and fits the logistic regression model. Section 3 discusses clearly the ML convergence problem. Application on modified score function in logistic regression model will discuss in Section 4 and it illustrates special case of modified function to give two equations that are used to estimate the parameters. Section 5 investigates the asymptotic properties for logistic regression model with making comparison between estimated parameters with ML method and reduction technique by simulated data. The discussion, conclusion and some general remarks about the results are in Section 6.

2. The Logistic Regression Model

The goal of a logistic regression analysis is to find the best fitting model to describe the relationship between an outcome and covariates where the outcome is dichotomous. [6] considered the logistic regression model is a member of the class of the generalized linear models. For more details of logistic model see [7] [8] [9] also [10] [11] [12].

The Model

Suppose now $y_i \sim \text{binomial}(m_i, \pi_i)$ where $y_i, i = 1, \dots, n$ is a response variable. Suppose that $y_i \in \{0, 1, \dots, m_i\}$ and π_i are related to a collection of covariates $(x_{i1}, x_{i2}, \dots, x_{ip})$ according to the equation

$$\log\left(\frac{\pi_i}{1 - \pi_i}\right) = \sum_{j=1}^p \beta_j x_{ij} = \beta^T x_i \quad (1)$$

We consider the special case $m_i = 1$ so $y_i \sim \text{binomial}(1, \pi_i)$ where π_i is the probability of success for each $i = 1, \dots, n$. We also define $\eta_i = \beta^T x_i$ so that

$$g(\pi_i) = \log\left(\frac{\pi_i}{1 - \pi_i}\right) = \eta_i \quad (2)$$

and

$$\pi_i = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)} \quad (3)$$

Here $g(\cdot)$ is called the logit link function and $\eta_i = \sum_{j=1}^p \beta_j x_{ij} = \beta^T x_i$ is the linear predictor.

There are some other link functions which can also be used, instead of the logit link function such as the probit link function

$$\eta = \Phi^{-1}(\pi) \leftrightarrow \pi = \Phi(\eta)$$

and the complementary log-log link function

$$\eta = \log(-\log(1 - \pi)) \leftrightarrow \pi = 1 - \exp(-e^\eta)$$

Fitting The Model

The logistic model when $y_i \sim \text{binomial}(m_i, \pi_i)$ with $m_i = 1$ can be fitted using the method of maximum likelihood to estimate the parameters. The first step is to construct the likelihood function which is a function of the unknown parameters. we choose those values of the parameters that maximize this function. The probability function of the model is

$$f(y_i) = \pi_i^{y_i} (1 - \pi_i)^{1-y_i} \tag{4}$$

where the likelihood function is

$$L(\pi_i | y_i) = \prod_{i=1}^n f(y_i | \pi_i) \tag{5}$$

Since the observations are independent, the likelihood function is as follows:

$$L(\pi_i | y_i) = \prod_{i=1}^n \pi_i^{y_i} (1 - \pi_i)^{1-y_i} \tag{6}$$

The maximum likelihood estimate of β is the value which maximizes the likelihood function. In general the log likelihood function is easier to work with mathematically and is:

$$l = \sum_{i=1}^n [y_i \log(\pi_i) + (1 - y_i) \log(1 - \pi_i)] \tag{7}$$

2.1. Special Case of the Logistic Model with Two Covariates

In this case the logistic regression model with two covariates, thus, $p = 2$, with one the general mean. So, we have β_0 and β_1 , such that

$$g(\pi_i) = \eta_i = \beta_0 + \beta_1 x_i \tag{8}$$

where x_i is now a scalar covariate and

$$\pi_i = \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)} \tag{9}$$

Therefore we can write the log-likelihood function as:

$$l(\beta_0, \beta_1) = \sum_{i=1}^n y_i (\beta_0 + \beta_1 x_i) - \log [1 + \exp(\beta_0 + \beta_1 x_i)] \tag{10}$$

To estimate the values of β_0 and β_1 we differentiate $l(\beta_0, \beta_1)$ in terms of β_0 and β_1 respectively as:

$$\frac{\partial l}{\partial \beta_0} = \sum_{i=1}^n y_i - \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)} = \sum_{i=1}^n (y_i - \pi_i) \tag{11}$$

$$\frac{\partial l}{\partial \beta_1} = \sum_{i=1}^n y_i x_i - \frac{x_i (\exp(\beta_0 + \beta_1 x_i))}{1 + \exp(\beta_0 + \beta_1 x_i)} = \sum_{i=1}^n (y_i - \pi_i) x_i \tag{12}$$

Now we set $\frac{\partial l}{\partial \beta_0} = 0$ and $\frac{\partial l}{\partial \beta_1} = 0$ and so the maximum likelihood estimates of β_0 and β_1 are the solution of the following equations

$$\sum_{i=1}^n y_i = \sum_{i=1}^n \pi_i \tag{13}$$

and

$$\sum_{i=1}^n y_i x_i = \sum_{i=1}^n \pi_i x_i \tag{14}$$

and will be denoted as $\hat{\beta}_0$ and $\hat{\beta}_1$. We know that for the logistic regression the last two equations are non linear in β_0 and β_1 , and we need to use a numerical method for their solution, such as Newton-Raphson method.

2.2. The Asymptotic Distribution of the (MLE)

The estimated parameters $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)'$, have an asymptotic distribution which is given by $\hat{\beta} \sim N(\beta, I(\beta)^{-1})$ where $I(\beta)$ is Fisher's information matrix defined as

$$I(\beta) = -E \begin{pmatrix} \frac{\partial^2 l}{\partial \beta_0^2} & \frac{\partial^2 l}{\partial \beta_0 \partial \beta_1} \\ \frac{\partial^2 l}{\partial \beta_1 \partial \beta_0} & \frac{\partial^2 l}{\partial \beta_1^2} \end{pmatrix} \tag{15}$$

where the matrix is evaluated at the MLE. For the logistic regression the estimated Fisher Information matrix can be written as

$$I(\hat{\beta}) = \begin{pmatrix} \sum_{i=1}^n \hat{\pi}_i (1 - \hat{\pi}_i) & \sum_{i=1}^n x_i \hat{\pi}_i (1 - \hat{\pi}_i) \\ \sum_{i=1}^n x_i \hat{\pi}_i (1 - \hat{\pi}_i) & \sum_{i=1}^n x_i^2 \hat{\pi}_i (1 - \hat{\pi}_i) \end{pmatrix} \tag{16}$$

where $\hat{\pi}_i = \frac{\exp(\hat{\eta})}{1 + \exp(\hat{\eta})}$ and $\hat{\eta} = \hat{\beta}_0 + \hat{\beta}_1 x_i$. The variance of $\hat{\beta}$ is approximated defined by $\text{Var}(\hat{\beta}) = I(\hat{\beta})^{-1}$.

3. Maximum Likelihood Convergence Problems

A problem occurs in estimating logistic regression models when the maximum likelihood estimates do not exist and one or more components of $\hat{\beta}$ are infinite. The one case of the occurrence of this problem is when all of the observations have the same response. For example, suppose that $m_i = 1$ and that all of the response variables equal zero *i.e.*, $\sum_{i=1}^n y_i = 0$. In this case the log-likelihood function is

$$l(\beta_0, \beta_1) = \sum_{i=1}^n -\log [1 + \exp(\beta_0 + \beta_1 x_i)] \tag{17}$$

Now differentiating $l(\beta_0, \beta_1)$ in terms of β_0 and β_1 respectively and setting equal to zero gives

$$\sum_{i=1}^n \pi_i = \sum_{i=1}^n \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 x_i)}{1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 x_i)} = 0 \quad (18)$$

and

$$\sum_{i=1}^n \pi_i x_i = \sum_{i=1}^n x_i \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 x_i)}{1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 x_i)} = 0 \quad (19)$$

The first equation has no solution because it is the sum of positive quantities and so cannot be equal to zero and satisfy the equation. To make this equation equal to zero we need to make β_0 larger and negative *i.e.* tend to $-\infty$. However, if precisely one of the response variable equal 1, the result maximum likelihood equation become

$$\sum_{i=1}^n \pi_i = 1 \quad (20)$$

$$\sum_{i=1}^n \pi_i x_i = x_1 \quad (21)$$

where we have assumed the numbers such that $y_1 = 1$. Here the maximum likelihood estimates is exist and the convergence of the MLE is achieved. Because the two previous equations are sum of positive quantities equal positive values. So as in first equation, if parameter is large and positive, then the sum is larger than one as well as if it is large and negative, then the sum is smaller than one and will not satisfy the equation, then we can find finite estimate of parameters which satisfy the equation.

4. Modified Score Function

Firth [2] proposed a method to reduce bias in MLE. The maximum likelihood convergence problem does not exist with the modified score function. The idea that extend and focus on two standard approaches have been extensively studied in the literature. The computationally-intensive jackknife method proposed by [13] [14]. The second approach simply substitutes $\hat{\beta}$ for the unknown β in $\frac{b(\beta)}{n}$. The point that discussed in case of small size sets of data, it is not uncommon that $\hat{\beta}$ is infinite in some samples of logistic regression models [15] [16]. We know that the maximum likelihood approach is dependent on the derivative the log-likelihood function as a solution to the score equation

$$\frac{\partial l(\beta)}{\partial \beta} = U(\beta) = 0 \quad (22)$$

[2] proposed that instead, we solve $U^*(\beta) = 0$, where the appropriate modification to $U(\beta)$ is:

$$U^*(\beta) = U(\beta) - I(\beta)b(\beta) \quad (23)$$

and the expected value of $\hat{\beta}$ proposed by [3], is given by:

$$E(\hat{\beta}) = \beta + b(\hat{\beta}) + O(n^{-1}) \tag{24}$$

where

$$b(\beta) = \frac{2E\left(\frac{\partial l}{\partial \beta} \frac{\partial l^2}{\partial \beta^2}\right) + E\left(\frac{\partial l^3}{\partial \beta^3}\right)}{2\left\{E\left(\frac{\partial l^2}{\partial \beta^2}\right)\right\}^2}$$

The variance of $\hat{\beta}$ is approximated defined by $\text{Var}(\hat{\beta}) \approx I(\hat{\beta})^{-1}$.

4.1. Modified Function with Logistic Regression Model

In this part we will apply the modified score function to simple logistic regression model. We know that the $O(n^{-1})$ bias vector given in the form

$b = (X^T W X)^{-1} X^T W \xi$ which proposed by [17]. Here $W \xi$ has i th element $h_i \left(\pi_i - \frac{1}{2}\right)$ and h_i is the i th diagonal element of the hat matrix

$$H = W^{1/2} X (X^T W X)^{-1} X^T W^{1/2}.$$

where $W = \text{diag}(\pi_i(1-\pi_i))$ and X is the design matrix. Then, the modified score function is written as

$$U^* = U - X^T W \xi \tag{25}$$

In this case, the modified score function $U^* = (U_0^*, U_1^*)$ gives two equations

$$U_0^* = \sum_{i=1}^n \left[\left(y_i + \frac{h_i}{2} \right) - (1 + h_i) \pi_i \right] = 0 \tag{26}$$

and

$$U_1^* = \sum_{i=1}^n \left[\left(y_i + \frac{h_i}{2} \right) - (1 + h_i) \pi_i \right] x_i = 0 \tag{27}$$

These are used to estimate the parameters.

4.2. Special Case of Modified Function

For more evaluation, we will discuss the behaviour of the adjusted score function when all the observation have the same response *i.e.* $\sum_{i=1}^n y_i = 0$. As a special case, suppose we have one explanatory variable x_i taking values 0 or 1. Before we calculate the adjusted score function, first calculate the form of h_i which we obtain from H . Here, h_i is the diagonal element of the H matrix and is

$$h_i = \frac{\pi_i(1-\pi_i)(X_2 - 2x_i X_1 + x_i^2 X_0)}{\Delta} \tag{28}$$

where $\Delta = X_0 X_2 - X_1^2$, $X_0 = n_0 \pi_0(1-\pi_0) + n_1 \pi_1(1-\pi_1)$, $X_1 = n_1 \pi_1(1-\pi_1)$ and $X_2 = n_1 \pi_1(1-\pi_1)$, where n_0 and n_1 are the number of observations of x equal to 0 and 1 respectively. Hence

$$h_0 = \frac{\pi_0(1-\pi_0)[n_1 \pi_1(1-\pi_1)]}{\Delta} \tag{29}$$

and

$$h_1 = \frac{\pi_1(1-\pi_1)[n_0\pi_0(1-\pi_0)]}{\Delta} \tag{30}$$

Therefore, when we set the adjusted score function $(U_0^*, U_1^*) = 0$ with $\sum_{i=1}^n y_i = 0$ we have

$$U_1^* = \sum_{i=1}^n \left[\frac{h_i}{2} - (1+h_i)\pi_i \right] x_i = 0 \tag{31}$$

This gives

$$\left[\frac{h_1}{2} - (1+h_1)\pi_1 \right] = 0 \tag{32}$$

and

$$\pi_1 = \frac{h_1}{2(1+h_1)} \tag{33}$$

Now,

$$U_0^* = \sum_{i=1}^n \left[\frac{h_i}{2} - (1+h_i)\pi_i \right] = 0 \tag{34}$$

and so

$$\left[h_1 \frac{n_1}{2} - n_1(1+h_1)\pi_1 \right] + \left[h_0 \frac{n_0}{2} - n_0(1+h_0)\pi_0 \right] = 0 \tag{35}$$

we get

$$\pi_0 = \frac{h_0}{2(1+h_0)} \tag{36}$$

Before calculate π_0 and π_1 we can consider the following way to calculate h_0 and h_1 . Let $A = n_1\pi_1(1-\pi_1)$ and $B = n_0\pi_0(1-\pi_0)$. Then, $X_0 = A + B$, $X_1 = A$ and $X_2 = A$, so, we can write Δ as

$$\Delta = X_0X_2 - X_1^2 = A^2 + AB - A^2 = AB = n_1\pi_1(1-\pi_1)n_0\pi_0(1-\pi_0) \tag{37}$$

Therefore, h_0 and h_1 can be written as

$$h_0 = \frac{\pi_0(1-\pi_0)[n_1\pi_1(1-\pi_1)]}{[n_1\pi_1(1-\pi_1)]n_0\pi_0(1-\pi_0)} = \frac{1}{n_0} \tag{38}$$

and

$$h_1 = \frac{\pi_1(1-\pi_1)[n_0\pi_0(1-\pi_0)]}{[n_1\pi_1(1-\pi_1)]n_0\pi_0(1-\pi_0)} = \frac{1}{n_1} \tag{39}$$

Then, we obtain

$$\pi_0 = \frac{h_0}{2(1+h_0)} = \frac{1/n_0}{2(1+1/n_0)} = \frac{1}{2(n_0+1)} \tag{40}$$

and

$$\pi_1 = \frac{h_1}{2(1+h_1)} = \frac{1/n_1}{2(1+1/n_1)} = \frac{1}{2(n_1+1)} \tag{41}$$

As a result of this example with $x = 0,1$ when $\sum_{i=1}^n y = 0$, we can say that, the estimate of parameters are finite. The modified function works well and the problem of convergence does not exist.

5. Simulation Study

The follow discussion are the simulation plan and the designs used in generating the data to identify the effect of sample size and proportion of events (the percentage of $y = 1$ or $y = 0$) on estimation of parameters. We will examine the precision of the estimation by calculating the variance of parameters obtained by simulation for the two approaches, MLE and Firth, and compare those with $I(\beta)^{-1}$ evaluated at the known values of β . The simulation study is designed as follows:

- 1) Three sample sizes have been used $n = 40$, $n = 120$ and $n = 500$.
- 2) For each sample size we choose x_i as a draw from $N(0,1)$. The x variables are fixed at these values throughout the simulation.
- 3) We choose β_0 and β_1 to give three cases. Choose $\beta_1 = 0.2$ and adjust β_0 so that over the covariates $\text{pr}(y = 1)$ is approximately (a) 0.5, (b) 0.1, (c) 0.05.
- 4) For each sample size and set of parameter values we perform 100,000 simulation.
- 5) Two approaches are used to estimate the parameters, MLE and the bias-reduced estimator Firth.

5.1. Results and Discussion of Sample Size $n = 500$

The simulation reported the accuracy of the estimation of $\text{Var}(\hat{\beta})$ using the information matrix. We calculate $\text{Var}(\hat{\beta}_0)$ and $\text{Var}(\hat{\beta}_1)$ for the simulated values of $\hat{\beta}_0$, $\hat{\beta}_1$ and also by evaluating $I(\beta)$ at the known values of β . The results in the **Table 1**, which shows the three cases of the proportion of $y = 1$, achieved the convergence of likelihood maximization algorithm.

As can be seen in **Table 1**, $\text{Var}\hat{\beta}_L$ Sim and $\text{Var}\hat{\beta}_F$ Sim are the variance of the parameters estimated by MLE and Firth method respectively. Ratio L and Ratio F denote the ratio of the variance estimated by MLE and Firth's method, respectively. The results showed that, both the variance of the parameters calculated from the simulation and the variance calculated by evaluating the information matrix at the known values of β are almost the same. We note that the ratio in the first case when $\text{pr}(y = 1)$ is 0.5 appeared nearly close to one but in the second case and the last case the ratio appeared slightly larger than in the first case.

The variance of parameters calculated by Firth's method were smaller than when calculated by MLE and the ratio in general was close to 1. Moreover, the bias ($\hat{\beta}_F - \beta$) was smaller.

5.2. Results and Discussion of Sample Size $n = 120$

In this part using the same way used in the previous case when $n = 500$. The

Table 1. Results of 100,000 simulations with sample size $n = 500$ and $(0.5, 0.1, 0.05)$ proportion of $y = 1$.

(a)

$\text{pr}(y = 1)$	parameter	$\text{Var}\hat{\beta}_L$ Sim	$\text{Var}\hat{\beta}_F$ Sim	$\text{Var}\hat{\beta}$ Inf	Ratio L	Ratio F
0.5	$\hat{\beta}_0$	0.00804	0.00813	0.00809	0.99	1.005
-	$\hat{\beta}_1$	0.00869	0.00862	0.00858	1.01	1.04
0.1	$\hat{\beta}_0$	0.02407	0.02359	0.02312	1.04	1.02
-	$\hat{\beta}_1$	0.02455	0.02439	0.02390	1.03	1.02
0.05	$\hat{\beta}_0$	0.04938	0.04560	0.04411	1.12	1.03
-	$\hat{\beta}_1$	0.04725	0.04656	0.04525	1.04	1.03

The variance of the parameters estimated by MLE and Firth with $(0.5, 0.1, 0.05)$ proportion of $y = 1$.

(b)

$\text{pr}(y = 1)$	parameter	$\hat{\beta}_L - \beta$	$\hat{\beta}_F - \beta$
0.5	$\hat{\beta}_0$	-0.0005	0.0003
-	$\hat{\beta}_1$	0.0007	-0.0001
0.1	$\hat{\beta}_0$	-0.0210	0.0003
-	$\hat{\beta}_1$	0.0004	0.0003
0.05	$\hat{\beta}_0$	-0.0400	-0.0011
-	$\hat{\beta}_1$	0.0026	0.0007

The bias value with $(0.5, 0.1, 0.05)$ proportion of $y = 1$.

results of simulation are shown in **Table 2**. Maximum likelihood convergence problems occurred (when $\text{pr}(y = 1) = 0.05$). Note that, there are many situations in which the likelihood function has no maximum, in which case we say that the maximum likelihood estimate does not exist. Consider the simulation which generating the data set 100,000 times, in some cases the coefficients reach to infinite in the final iterations and so, we have not results of the estimation $\hat{\beta}_0$ and $\hat{\beta}_1$, that result in at which point the algorithm has not converged. In our simulation we consider the cases that not achieved the converges algorithm.

Here for only 99,806 (99%) of the data sets was it possible to obtain finite estimates of β_0 and β_1 converged. Moreover, the variance of the parameters $\hat{\beta}_0$ and $\hat{\beta}_1$ is large. This is because even though convergence is achieved when $\sum_{i=1}^n y_i = 1$, There are some very large negative values of $\hat{\beta}$. In the other two cases of $\text{pr}(y = 1) = (0.5, 0.1)$ we achieved ML convergence in every simulation. We note that the ratio is nearly one but is a bit high when compared with case of $n = 500$. Firth's approach showed reasonable results, all cases achieved the maximum likelihood convergence. Moreover, the ratio was better than MLE approach as well as the bias $\hat{\beta}_F - \beta$.

5.3. Results and Discussion of Sample Size $n = 40$

We used the same analysis as in the previous cases with $n = 40$. As can be seen in **Table 3**, the results showed that, MLE approach had convergence problems,

Table 2. Results of 100,000 simulations with sample size $n = 120$ and (0.5, 0.1, 0.05) proportion of $y = 1$.

(a)

$pr(y = 1)$	parameter	$Var\hat{\beta}_l$ Sim	$Var\hat{\beta}_f$ Sim	Var Inf	Ratio L	Ratio F
0.5	$\hat{\beta}_0$	0.034	0.0335	0.0337	1.04	0.99
-	$\hat{\beta}_1$	0.035	0.0331	0.0326	1.07	1.01
0.1	$\hat{\beta}_0$	0.125	0.1037	0.0951	1.32	1.09
-	$\hat{\beta}_1$	0.105	0.0942	0.0885	1.27	1.06
0.05	$\hat{\beta}_0$	230.94	0.2377	0.1811	-	1.31
-	$\hat{\beta}_1$	37.96	0.2003	0.1669	-	1.20

The variance of the parameters estimated by MLE and Firth with (0.5, 0.1, 0.05) proportion of $y = 1$.

(b)

$pr(y = 1)$	parameter	$\hat{\beta}_l - \beta$	$\hat{\beta}_f - \beta$
0.5	$\hat{\beta}_0$	-0.0005	-0.0004
-	$\hat{\beta}_1$	0.0056	-0.0002
0.1	$\hat{\beta}_0$	-0.0845	-0.0021
-	$\hat{\beta}_1$	0.0055	0.0008
0.05	$\hat{\beta}_0$	-0.4480	-0.0120
-	$\hat{\beta}_1$	0.090	-0.0027

The bias value with (0.5, 0.1, 0.05) proportion of $y = 1$.

Table 3. Results of 100,000 simulations with sample size $n = 40$ and (0.5, 0.1, 0.05) proportion of $y = 1$.

(a)

$pr(y = 1)$	parameter	$Var\hat{\beta}_l$ Sim	$Var\hat{\beta}_f$ Sim	Var Inf	Ratio L	Ratio F
0.5	$\hat{\beta}_0$	0.12184	0.1076	0.10780	1.13	0.99
-	$\hat{\beta}_1$	0.14889	0.1222	0.11766	1.22	1.04
0.1	$\hat{\beta}_0$	215.38	0.3702	0.29139	-	1.27
-	$\hat{\beta}_1$	81.22	0.3962	0.34304	-	1.16
0.05	$\hat{\beta}_0$	820.47	0.5080	0.55184	-	0.92
-	$\hat{\beta}_1$	295.64	0.5936	0.65309	-	0.91

The variance of the parameters estimated by MLE and Firth with (0.5, 0.1, 0.05) proportion of $y = 1$.

(b)

$pr(y = 1)$	parameter	$\hat{\beta}_l - \beta$	$\hat{\beta}_f - \beta$
0.5	$\hat{\beta}_0$	0.0029	-0.0001
-	$\hat{\beta}_1$	0.0220	-0.0010
0.1	$\hat{\beta}_0$	-1.2000	-0.0150
-	$\hat{\beta}_1$	0.3830	-0.014
0.05	$\hat{\beta}_0$	-3.9900	0.0440
-	$\hat{\beta}_1$	1.288	-0.0940

The bias value with (0.5, 0.1, 0.05) proportion of $y = 1$.

98,273 (98%) and 85,967 (86%) of data sets achieved ML convergence when $\text{pr}(y=1)$ was (0.1, 0.05), respectively. Convergence was only achieved in every simulation in the case of $\text{pr}(y=1) = 0.5$, where the ratio was nearly close to one, but is a bit high from previous cases. Moreover, we found the same problem as discussed in the case of $n = 120$, in that the variance of the parameters $\hat{\beta}_0$ and $\hat{\beta}_1$ is large. However, when we use Firth's approach, all data sets achieved M.L convergence. Moreover, the ratio was better than M.L.E approach as well as the bias $\hat{\beta}_F - \beta$ being smaller.

6. Conclusion

Attention has been directed in this work to determine the behaviour of the asymptotic estimation of parameters by two methods—MLE and bias reduction technique compared with the result of the information matrix. In fact in regular convergence problem the modified score function appeared appropriate behaviour, which denoted that the bias form may be removed from the MLE by reduction bias term. The asymptotic variance of the MLE may be appeared as strange behaviour, and the results shown variance of the parameters were large in some cases, even though convergence is achieved. It is denoted that there are some very large negative values of $\hat{\beta}$, as shown in results section. We can report that the small sample size and the value of $\text{pr}(y=1)$ have an effect on behaviour estimation of parameters when using MLE. Clearly, we found convergence problem for some combinations of sample size and $\text{pr}(y=1)$. The approach of Firth appeared a moderate results that the data sets in all cases of sample size and $\text{pr}(y=1)$ achieved ML convergence. Overall, we can consider the bias reduction technique is worked well and has a moderate behaviour almost with all cases which have been investigated. Moreover, the convergence problem is not only effective on behaviour of the MLE, and although the convergence is achieved, the variance of the parameters estimates appeared large value.

References

- [1] Cox, D.R. and Hinkley, D.V. (1974) *Theoretical Statistics*. Chapman and Hall, London.
- [2] Firth, D. (1993) Bias Reduction of Maximum Likelihood Estimates. *Biometrika*, **80**, 27-38. <https://doi.org/10.1093/biomet/80.1.27>
- [3] Anderson, J.A. and Richardson, C. (1979) Logistic Discrimination and Bias Correction in Maximum Likelihood Estimation. *Technometrics*, **21**, 71-78.
- [4] McCullagh, P. (1986) The Conditional Distribution of Goodness-of-Fit Statistics for Discrete Data. *Journal of the American Statistical Association*, **81**, 104-107. <https://doi.org/10.1080/01621459.1986.10478244>
- [5] Shenton, L.R. and Bowman, K.O. (1977) Maximum Likelihood Estimation in Small Samples. *Griffin's Statistical Monograph* No. 38, London.
- [6] Nelder, J.A. and Wedderburn, R.W.M. (1972) Generalized Linear Models. *Journal of the Royal Statistical Society, Series A*, **135**, 370-384. <https://doi.org/10.2307/2344614>

- [7] Dobson, A. (1990) An Introduction to Generalized Linear Models. Chapman and Hall, London.
- [8] Dobson, A.J. and Barnett, A.G. (2008) An Introduction to Generalized Linear Models. 3rd Edition, Chapman and Hall, New York.
- [9] Kleinbaum, D.G. (1994) Logistic Regression: A Self-Learning Text. Springer-Verlag, New York. <https://doi.org/10.1007/978-1-4757-4108-7>
- [10] Hilbe, J.M. (2009) Logistic Regression Model. Chapman and Hall, New York.
- [11] Hosmer, D.W. and Lemeshow, S. (2000) Applied Logistic Regression. Wiley, Chichester. <https://doi.org/10.1002/0471722146>
- [12] Hosmer, D., Lemeshow, S. and Sturdivant, R.X. (2013) Applied Logistic Regression. 3rd Edition, Wiley, Chichester. <https://doi.org/10.1002/9781118548387>
- [13] Quenouille, M.H. (1949) Approximate Tests of Correlation in Time-Series. *Journal of the Royal Statistical Society: Series B*, **11**, 68-84.
- [14] Quenouille, M.H. (1956) Notes on Bias in Estimation. *Biometrika*, **43**, 353-360. <https://doi.org/10.1093/biomet/43.3-4.353>
- [15] Albert, A. and Anderson, J.A. (1984) On the Existence of Maximum Likelihood Estimates in Logistic Regression Models. *Biometrika*, **71**, 1-10. <https://doi.org/10.1093/biomet/71.1.1>
- [16] Clogg, C.C., Rubin, D.B., Schenker, N., Schultz, B. and Weidman, L. (1991) Multiple Imputation of Industry and Occupation Codes in Census Public-Use Samples Using Bayesian Logistic Regression. *Journal of the American Statistical Association*, **86**, 68-78. <https://doi.org/10.1080/01621459.1991.10475005>
- [17] McCullagh, P. and Nelder, J.A. (1989) Linear Models. 2nd Edition, Chapman and Hall, London.



Open Access Library

Submit or recommend next manuscript to OALib Journal and we will provide best service for you:

- Publication frequency: Monthly
- 9 [subject areas](#) of science, technology and medicine
- Fair and rigorous peer-review system
- Fast publication process
- Article promotion in various social networking sites (LinkedIn, Facebook, Twitter, etc.)
- Maximum dissemination of your research work

Submit Your Paper Online: [Click Here to Submit](#)

Or Contact service@oalib.com