Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF
AGRICULTURAL STATISTICS 70(3) 2016 227–235

Improved Estimation in Logistic Regression through Quadratic Bootstrap Approach: An Application in Agricultural Ergonomics

Arpan Bhowmik[1], Ramasubramanian V.[2], Anil Rai[1], Adarsh Kumar[3] and Madan Gopal Kundu[4]
[1]ICAR- Indian Agricultural Statistics Research Institute, New Delhi
[2]ICAR- Central Institute of Fisheries Education, Mumbai
[3]ICAR-Indian Agricultural Research Institute, New Delhi
[4]Novartis Pharmaceutical Corporations, East Hanover, New Jersey–07876, USA

Received 11 December 2015; Revised 02 September 2016; Accepted 15 September 2016

SUMMARY

In this paper, quadratic bootstrap estimation procedure for improved estimation under logistic regression set up with one regressor based on Claeskens *et al*. (2003) for classification of data pertaining to the area of agricultural ergonomics have been discussed. Here, presence or absence of discomfort for the farm labourers in operating farm machineries has been considered as the dependent variable. A comparison in terms of confidence interval and classificatory ability of the logistic regression model between the usual maximum likelihood estimator and the quadratic bootstrap based estimator have been made based on real experimental situation in the field of agricultural ergonomics. The performance of quadratic bootstrap based estimator has been found to be better both in terms of length of the confidence interval of the parameter and classificatory ability of the model. Further, a bias corrected estimator based on quadratic bootstrap estimator following Claeskens *et al.* (2003) has also been obtained. A simulation study has been carried out which illustrates the improvement of bias corrected estimation over the usual maximum likelihood approach in terms of mean square error of the estimators and efficiency factor.

Keywords: Bias correction, Classificatory power, Confidence interval, Ergonomics, Quadratic bootstrap, Simulation.

1. INTRODUCTION

The most preferred model for analysis of binary (dichotomous) responses is the logistic regression model. By considering the two distinct responses as two groups, the usual logistic regression model can thus be reformulated as a classification technique in the lines of discriminant function analysis. Logistic regression have found profound applications in different fields such as agriculture, medical science (epidemiology and health), psychology etc. (Johnson *et al.* (1996), Tsien *et al.* (1998), Misra *et al.* (2002), Gent *et al.* (2003), Mila *et al.* (2004), Ayan and Garcia (2008) etc.). However, work on logistic regression in the field of ergonomics seems to be limited in literature. Vergara and Page (2002) classified lumbar discomfort/absence of discomfort by relating with back posture and mobility in sitting-posture using both discriminant analysis and logistic regression. Bhowmik *et al*. (2011) employed logistic regression model for classification of presence or absence of discomfort for the farm labourers in operating farm machineries with a set of quantitative and qualitative regressor variables. They have identified a single best regressor by employing variable selection based on collinearity diagnostics and stepwise logistic regression. Further, comparisons were made between the performances of logistic regression models with that of the discriminant function analysis.

We know that, maximum likelihood estimation (MLE) approach is most used approach to estimate the parameters under logistic regression. However, in recent days, bootstrap based methods have

Corresponding author: Arpan Bhowmik
E-mail address: arpan.stat@gmail.com

found wide applications in the area of regression analysis, particularly in nonlinear models. Over time there arose a spate of modifications in the bootstrap method. In case of ordinary bootstrap method, several independent samples are randomly drawn with replacement from the original sample with sizes same as that of the original sample and the estimates of any population parameter is computed from the estimates obtained from these resamples. It is well-known that the bootstrap can be very computer-intensive, especially if no analytic method can be used and simulation based approximations are required. Work related to the use of bootstrapping are large in number however those under logistic regression setting seems to be very few (for example Swapnepoel and Frangos 1994, Lee 1990, Claeskens and Aerts 2000, Aerts and Claeskens 2001 etc.). Claeskens *et al*. (2003) moved one step further by proposing a quadratic bootstrap method of estimation in logistic regression which had their underprintings from Claeskens and Aerts (2000). A bias corrected estimator based on the quadratic bootstrap estimator has also been constructed by them.

In this paper, based on a real experimental setup in the field of Agricultural Ergonomics, we have obtained the quadratic bootstrap estimator under logistic regression model following Claeskens *et al*. (2003) based on a single explanatory variable viz. load given to the farm machinery during farm operation. This single explanatory variable has been considered since Bhowmik *et al*. (2011) employed logistic regression modeling in agricultural ergonomics and have identified the single best regressor 'load given to the farm machinery during farm operation' for classifying presence or absence of discomfort. It has been emphasized here that the same best regressor has been considered in the present study also albeit for generating larger dataset using simulation and applying improved logistic regression model via quadratic bootstrap. Therefore, the performances in terms of length of the confidence interval of the model parameter using the quadratic bootstrap estimator based on the original sample have been compared with those of the usual maximum likelihood estimates of the original sample. The classificatory ability of logistic regression model fitted based on both these estimates of the parameter obtained from the original sample has also been compared. Further, on the lines of Claeskens *et al.* (2003), a bias corrected estimator under the logistic regression model with single explanatory variable has also been obtained through quadratic bootstrap estimates via simulation. The performance of bias corrected estimator has been compared with the usual maximum likelihood estimator through a simulation study by means of Mean Square error (MSE) and efficiency factor.

2. ESTIMATION IN LOGISTIC REGRESSION MODEL

Estimation and testing are two important aspects of regression analysis. The usual method of estimation under logistic regression is Maximum Likelihood Estimation method (MLE). Another method called quadratic bootstrap improves the estimation by giving a bias corrected estimator. The mean squared error obtained for such an estimator yields a narrower confidence interval. The procedure of employing the quadratic bootstrap method of estimation in logistic regression as in Claeskens *et al*. (2003) is discussed below.

2.1 Maximum Likelihood Estimation in Logistic Regression

Let, X represent an explanatory variable for which there are p levels. Let, all the p levels of the explanatory variable to be considered as p different populations. Let $Y_{i1}, \ldots, Y_{in_i}$ be independent identically distributed Bernoulli random variables with probability function $f_i(y) = \pi_i^{y_{ij}} (1 - \pi_i)^{1-y_{ij}}$, $y = 0, 1$; $i = 1, 2, \ldots, p$; $j = 1, \ldots, n_i$ where, Y_{ij} indicate whether the j^{th} outcome in population i is a "success" or not, π_i indicates the probability of success. Let n_i be the number of replications at x_i.

The number of levels p is considered as fixed whereas the observation number n_i of distinct populations become large as total number of observations $n = \sum_{i=1}^{p} n_i$ tends to infinity.

Logistic regression implies the success probability should be modeled as a function of covariates through model parameters $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_r)$.

For example, in a logistic model (r = 1)

$$P(Y_{ij} = 1) = \pi_i = \frac{1}{1 + \exp[-(\beta_0 + \beta_1 X_1)]}$$

$$i = 1, \ldots, p, \; j = 1, \ldots, n_i \qquad (1)$$

Now, for a logistic regression model with parameters $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_r)$, maximum likelihood estimates for associated populations is based on (r + 1) dimensional score functions (including the intercept) as

$$\psi_i(Y_{ij}, \hat{\boldsymbol{\beta}}_n) = (\partial / \partial\hat{\boldsymbol{\beta}}) \log f_i(y, \boldsymbol{\beta}), \; i = 1, \ldots, p \qquad (2)$$

More specifically, the score vector for logistic regression with r =1 i.e. one regressor is given as follows:

$$\psi_i\left(\mathrm{Y}_{ij}, \hat{\boldsymbol{\beta}}_n\right) = \left[\psi_{i1}\left(Y_{ij}, \hat{\boldsymbol{\beta}}_n\right) \psi_{i2}\left(Y_{ij}, \hat{\boldsymbol{\beta}}_n\right)\right]'$$

$$= \left[y_{ij} - \frac{1}{1 + \exp\left\{-\left(\hat{\boldsymbol{\beta}}_0 + \hat{\boldsymbol{\beta}}_1 x_{\mathrm{i}}\right)\right\}} y_{ij} \quad - \frac{x_i}{1 + \exp\left\{-\left(\hat{\boldsymbol{\beta}}_0 + \hat{\boldsymbol{\beta}}_1 x_{\mathrm{i}}\right)\right\}} \right]' \qquad (3)$$

Solving the system of equations

$$\sum_{i=1}^{p} \sum_{j=1}^{n_i} \psi_i\left(Y_{ij}, \boldsymbol{\beta}\right) = 0 \qquad (4)$$

leads to the maximum likelihood estimator $\hat{\boldsymbol{\beta}}_n$ for $\boldsymbol{\beta}$.

2.2 Quadratic Bootstrap and Bias Corrected Estimator

The quadratic bootstrap procedure can be used for finite sample bias correction. Let, a large number, say B, resamples are taken, resulting in a set of B quadratic bootstrap estimators $\hat{\boldsymbol{\beta}}_n^{*1}, \ldots, \hat{\boldsymbol{\beta}}_n^{*\mathrm{B}}$. From this set of quadratic bootstrap estimator, a bias corrected estimator can be obtained as

$$\hat{\boldsymbol{\beta}}_n^{bc} = 2\hat{\boldsymbol{\beta}}_n - \frac{1}{B} \sum_{u=1}^{B} \hat{\boldsymbol{\beta}}_n^{*u} \qquad (5)$$

here the quadratic bootstrap estimator $\hat{\boldsymbol{\beta}}^{*u}$ is defined as:

$$\hat{\boldsymbol{\beta}}_n^{*u} = \hat{\boldsymbol{\beta}}_{\mathrm{v}} + U_n^* - \frac{1}{2}\left(\sum_{i=1}^{p} \sum_{j=1}^{n_i} \tilde{\psi}_{ij}^*\left(\hat{\boldsymbol{\beta}}_n\right)\right)^{-1} \left(\sum_{k=0}^{r} \sum_{l=0}^{r} \sum_{i=1}^{p} \sum_{j=1}^{n_i} \tilde{\tilde{\psi}}_{\mathrm{ij}}^*\left(\hat{\boldsymbol{\beta}}_n\right)_{k,\mathrm{l}} \mathrm{U}_{\mathrm{nk}}^* \mathrm{U}_{\mathrm{nl}}^*\right) \qquad (6)$$

for all i = 1, …, p, j = 1, …, n_i and u = 1, 2, ... B with

$$U_n^* = -\left(\sum_{i=1}^{p} \sum_{j=1}^{n_i} \tilde{\psi}_{ij}^*\left(\hat{\boldsymbol{\beta}}_n\right)\right)^{-1} \left(\sum_{i=1}^{p} \sum_{j=1}^{n_i} \psi_{ij}^*\left(\hat{\boldsymbol{\beta}}_n\right)\right).$$

The quadratic bootstrap estimators are based on the values $\left(\psi_{ij}^*\left(\hat{\boldsymbol{\beta}}_n\right), \tilde{\psi}_{ij}^*\left(\hat{\boldsymbol{\beta}}_n\right), \tilde{\tilde{\psi}}_{ij}^*\left(\hat{\boldsymbol{\beta}}_n\right)\right)$, j = 1, …, n_i and i = 1, 2 … p taken with replacement from the set

$$\left\{\left(\psi_i\left(Y_{ij}, \hat{\boldsymbol{\beta}}_n\right), \left(\partial / \partial\boldsymbol{\beta}\right) \psi_i\left(Y_{ij}, \hat{\boldsymbol{\beta}}_n\right), \left(\partial^2 / \partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^{\mathrm{T}}\right) \psi_i\left(Y_{ij}, \hat{n}_n\right)\right), \; j = 1, \ldots, n_{\mathrm{i}}, i = 1, \ldots, p\right\}$$

The first term $\hat{\beta}_n$ at the right-hand side of Equation (6) is the maximum likelihood estimator of the original sample obtained by solving Equation (4). It is expected that, the representation of the random variation about estimate $\hat{\boldsymbol{\beta}}_n^{*u}$ would be improved through the last term at the right-hand side of Equation (6). Here $\psi_i\left(Y_{ij}, \hat{\boldsymbol{\beta}}_n\right)$ is a 2 × 1 vector defined in Equation (3). The 2 × 2 matrix of the first derivatives of the score function i.e. $\left(\partial / \partial\boldsymbol{\beta}\right) \psi_{\mathrm{i}}\left(\mathrm{Y}_{\mathrm{ij}}, \hat{\boldsymbol{\beta}}_{\mathrm{n}}\right)$ is given by

$$\left(\partial / \partial\boldsymbol{\beta}\right) \psi_{\mathrm{i}}\left(\mathrm{Y}_{\mathrm{ij}}, \hat{\boldsymbol{\beta}}_{\mathrm{n}}\right) = \begin{bmatrix} \left(\partial / \partial\beta_0\right) \psi_{\mathrm{i}1}\left(\mathrm{Y}_{\mathrm{ij}}, \hat{\boldsymbol{\beta}}_{\mathbf{n}}\right) & \left(\partial / \partial\beta_1\right) \psi_{i1}\left(Y_{\mathrm{ij}}, \hat{\boldsymbol{\beta}}_{\mathbf{n}}\right) \\ \left(\partial / \partial\beta_0\right) \psi_{\mathrm{i}2}\left(\mathrm{Y}_{ij}, \hat{\boldsymbol{\beta}}_{\mathbf{n}}\right) & \left(\partial / \partial\beta_1\right) \psi_{\mathrm{i}2}\left(Y_{\mathrm{ij}}, \hat{\boldsymbol{\beta}}_n\right) \end{bmatrix}_{\boldsymbol{\beta}_0 = \hat{\boldsymbol{\beta}}_0, \boldsymbol{\beta}_1 = \hat{\boldsymbol{\beta}}_0}$$

$$= \begin{bmatrix} \dfrac{-\exp\left\{-\left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right)\right\}}{\left(1+\exp\left\{-\left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right)\right\}\right)^2} & \dfrac{-x_i \exp\left\{-\left(\hat{n}_0 + \hat{n}_1 x_i\right)\right\}}{\left(1+\exp\left\{-\left(\hat{n}_0 + \hat{n}_1 x_i\right)\right\}\right)^2} \\ \dfrac{-x_i \exp\left\{-\left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right)\right\}}{\left(1+\exp\left\{-\left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right)\right\}\right)^2} & \dfrac{-x_i^2 \exp\left\{-\left(\hat{n}_0 + \hat{n}_1 x_i\right)\right\}}{\left(1+\exp\left\{-\left(\hat{n}_0 + \hat{n}_1 x_i\right)\right\}\right)^2} \end{bmatrix}$$

The second partial derivatives of $\psi_i\left(Y_{ij}, \hat{\beta}_n\right)$ are taken w.r.t $(k, l)^{th}$ components of β, here k, l = 0,1 and the resulting vectors are of dimension 2 × 1. The corresponding vectors are given as follows.

The second partial derivative of $\psi_i\left(Y_{ij}, \hat{\beta}_n\right)$ with respect to $(0, 0)^{th}$ component is:

$$\begin{bmatrix} \left(\partial^2 / \partial\beta_0^2\right) \psi_{i1}\left(Y_{ij}, \hat{\beta}_n\right) \\ \left(\partial^2 / \partial\beta_0^2\right) \psi_{i2}\left(Y_{ij}, \hat{\beta}_n\right) \end{bmatrix}_{\beta_0 = \hat{\beta}_0, \beta_1 = \hat{\beta}_1}$$

$$= \begin{bmatrix} \dfrac{\exp\left\{-\left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right)\right\} - \exp\left\{-2\hat{\beta}_0 - 2\hat{\beta}_1 x_i\right\}}{\left(1+\exp\left\{-\left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right)\right\}\right)^3} \\ \dfrac{x_i \exp\left\{-\left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right)\right\} - x_i \exp\left\{-2\hat{\beta}_0 - 2\hat{\beta}_1 x_i\right\}}{\left(1+\exp\left\{-\left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right)\right\}\right)^3} \end{bmatrix}$$

The second partial derivative of $\psi_i\left(Y_{ij}, \hat{\beta}_n\right)$ with respect to $(0, 1)^{th}$ component is:

$$\begin{bmatrix} \left(\partial^2 / \partial\beta_0 \partial\beta_1\right) \psi_{i1}\left(Y_{ij}, \hat{\beta}_n\right) \\ \left(\partial^2 / \partial\beta_0 \partial\beta_1\right) \psi_{i2}\left(Y_{ij}, \hat{\beta}_n\right) \end{bmatrix}_{\beta_0 = \hat{\beta}_0, \beta_1 = \hat{\beta}_1}$$

$$= \begin{bmatrix} \dfrac{x_i \exp\left\{-\left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right)\right\} - x_i \exp\left\{-2\hat{\beta}_0 - 2\hat{\beta}_1 x_i\right\}}{\left(1+\exp\left\{-\left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right)\right\}\right)^3} \\ \dfrac{x_i^2 \exp\left\{-\left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right)\right\} - x_i^2 \exp\left\{-2\hat{\beta}_0 - 2\hat{\beta}_1 x_i\right\}}{\left(1+\exp\left\{-\left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right)\right\}\right)^3} \end{bmatrix}$$

The second partial derivative of $\psi_i\left(Y_{ij}, \hat{\beta}_n\right)$ with respect to $(1, 0)^{th}$ component is:

$$\begin{bmatrix} \left(\partial^2 / \partial\beta_1 \partial\beta_0\right) \psi_{i1}\left(Y_{ij}, \hat{\beta}_n\right) \\ \left(\partial^2 / \partial\beta_1 \partial\beta_0\right) \psi_{i2}\left(Y_{ij}, \hat{\beta}_n\right) \end{bmatrix}_{\beta_0 = \hat{\beta}_0, \beta_1 = \hat{\beta}_1}$$

$$= \begin{bmatrix} \dfrac{x_i \exp\left\{-\left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right)\right\} - x_i \exp\left\{-2\hat{\beta}_0 - 2\hat{\beta}_1 x_i\right\}}{\left(1+\exp\left\{-\left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right)\right\}\right)^3} \\ \dfrac{x_i^2 \exp\left\{-\left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right)\right\} - x_i^2 \exp\left\{-2\hat{\beta}_0 - 2\hat{\beta}_1 x_i\right\}}{\left(1+\exp\left\{-\left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right)\right\}\right)^3} \end{bmatrix}$$

The second partial derivative of $\psi_i(Y_{ij}, \hat{\beta}_n)$ with respect to $(1, 1)^{th}$ component is:

$$\begin{bmatrix} \left(\partial^2 / \partial\beta_1^2\right) \psi_{i1}\left(Y_{ij}, \hat{\beta}_n\right) \\ \left(\partial^2 / \partial\beta_1^2\right) \psi_{i2}\left(Y_{ij}, \hat{\beta}_n\right) \end{bmatrix}_{\beta_0 = \hat{\beta}_0, \beta_1 = \hat{\beta}_1}$$

$$= \begin{bmatrix} \dfrac{x_i^2 \exp\left\{-\left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right)\right\} - x_i^2 \exp\left\{-2\hat{\beta}_0 - 2\hat{\beta}_1 x_i\right\}}{\left(1+\exp\left\{-\left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right)\right\}\right)^3} \\ \dfrac{x_i^3 \exp\left\{-\left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right)\right\} - x_i^3 \exp\left\{-2\hat{\beta}_0 - 2\hat{\beta}_1 x_i\right\}}{\left(1+\exp\left\{-\left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right)\right\}\right)^3} \end{bmatrix}$$

The estimated bias of the quadratic bootstrap estimator is calculated as

$$\frac{1}{B}\sum_{u=1}^{B} \hat{\beta}_n^{*u} - \hat{\beta}_n \tag{7}$$

For the present study, first we have obtained a good number of quadratic bootstrap samples from the original sample and hence obtain the quadratic bootstarp estimator of the model parameter with single explanatory variable along with the maximum likelihood estimator of the original sample. The length of the confidence interval for both the estimators of model parameter obtained through the original sample have been computed and comparison has been made. Later, a large number of simulated samples have also been generated from the original sample and for each of the simulated sample, both MLE and bias corrected estimates have been constructed. Finally comparison has been made between these two approaches on the basis of mean square error (MSE) and efficiency. The mean square errors for both the MLE and bias corrected estimator have been calculated as:

$$\text{MSE}\left(\hat{\beta}_n\right) = \text{variance}\left(\hat{\beta}_n\right) + \left[\text{bias}\left(\hat{\beta}_n\right)\right]^2 \quad (8)$$

$$\text{MSE}\left(\hat{\beta}_n^{bc}\right) = \text{variance}\left(\hat{\beta}_n^{bc}\right) + \left[\text{bias}\left(\hat{\beta}_n^{bc}\right)\right]^2 \quad (9)$$

3. SOURCE AND EXTENT OF DATA

Indian farm employs 225 million workers, constituting 10 percent of total world's workforce in agriculture activities (Ram *et al.* 2008). Many of the tasks are performed manually. They rely on rotary power input; it may be pedal, hand rocking or flywheel-operated machines. Human energy is used to operate different machines. Working environment of farm is labour intensive and strenuous. In the farm environment, machines are generally operated at a greater physiological cost and postural stress leading to discomfort and fatigue of farmers depending upon posture, force application, quantum and frequency. The farm labourers experience discomfort in hands and legs in general and thighs, knees, feet, legs, back, palms, buttocks etc. in particular. Researchers have categorized the overall discomfort of the farm labour on the basis of various factors like loads given to various farm machineries, heart beat per unit time, oxygen consumption per unit time, feeling of discomfort in various body parts during farm operation, % of aerobic capacity of the farm labour during operation, the mode of operation such as stepper, pedal, bicycle etc.

For the present study, the data has been taken in the area of agricultural ergonomics from the Division of Agricultural Engineering, ICAR-Indian Agricultural Research Institute (IARI), New Delhi. The variable considered as the dependent variable (*Y*) for the present study is the presence or absence of discomfort for the farm labourers during farm operation with two levels 0 and 1 depending upon whether discomfort is absent or present. The single explanatory variables (*X*) considered is the load given to farm machineries during farm operations. Here, the variable Load given to the farm machinery was having five levels viz. 0W(no load), 0.90W, 1.80W, 2.70W and 3.60W (W is the unit of power i.e. Watt). In total 135 observations were available for the study which were made upon nine subjects (farm labourers) over three independent time periods under each of the five levels of the loads given to farm machineries during farm operations. Out of the 135 observations available, 80% i.e. 108 observations have been selected randomly. These 108 observations consist of the original sample. The remaining 20% i.e. (135-108) = 27 observations have been considered as hold out data set for model validation in terms of classifying ability of the model.

4. QUADRATIC BOOTSTRAP ESTIMATES VS MAXIMUM LIKELIHOOD ESTIMATES

Using the original sample of size 108, 200 quadratic bootstrap estimates have been constructed using Equation (6). The mean of these 200 estimates has been calculated and the resulting value is taken as a single quadratic bootstrap estimate given by $\hat{\beta}_0^*$ = -5.78 and $\hat{\beta}_1^*$ = 4.36. Along with the quadratic bootstrap estimates, maximum likelihood estimates of the model parameters under logistic regression setup with single explanatory variable has also been obtained from the original sample as $\hat{\beta}_0$ = -4.60 and $\hat{\beta}_1$ = 3.46. Thereafter, for both the estimates, 95% confidence intervals have been calculated and are given in Table 1. Comparison has been made between the length of confidence interval of these two estimates. From Table 1, it has been found that the length of confidence interval for quadratic bootstrap estimates is shorter than that of maximum likelihood estimates. Thus there is a reduction in confidence interval for quadratic bootstrap estimates as compared to the MLE based estimates on the original sample. Thus quadratic bootstrap performs better than maximum likelihood estimates. Along with the comparison, the 200 quadratic bootstrap estimates are also grouped in 10 groups and frequency of each group has been calculated in order to draw histograms (given in Fig. 1 and Fig. 2) for knowing the distribution pattern of the parameters. From Fig. 1 and Fig. 2, it can be said that the distribution of the intercept parameter is negatively skewed whereas it is positively skewed for the slope parameter under the logistic regression model considered for the present study.

Table 1. 95% confidence intervals for maximum likelihood estimates and quadratic bootstrap estimates based on original sample

Estimation Method	Estimates	95% Confidence Interval	Length of Confidence Interval
MLE	=- 4.60	(- 6.56, - 2.65)	3.90
	= 3.46	(2.08, 4.85)	2.77
Quadratic Bootstrap	= - 5.78	(- 7.18, - 4.33)	2.85
	= 4.36	(3.48, 5.25)	1.77

Table 2. Comparison between classificatory abilities of logistic regression models fitted based on maximum likelihood estimates and quadratic bootstrap estimates based on the original sample

Method of Estimation	Correct Classificatioraratio	Sensitivity	Specificity	False Positive Rate	False Negative Rate
Maximum likelihood	92.59	94.12	90.00	5.88	10.00
Quadratic bootstrap	93.10	94.75	91.20	5.01	9.21

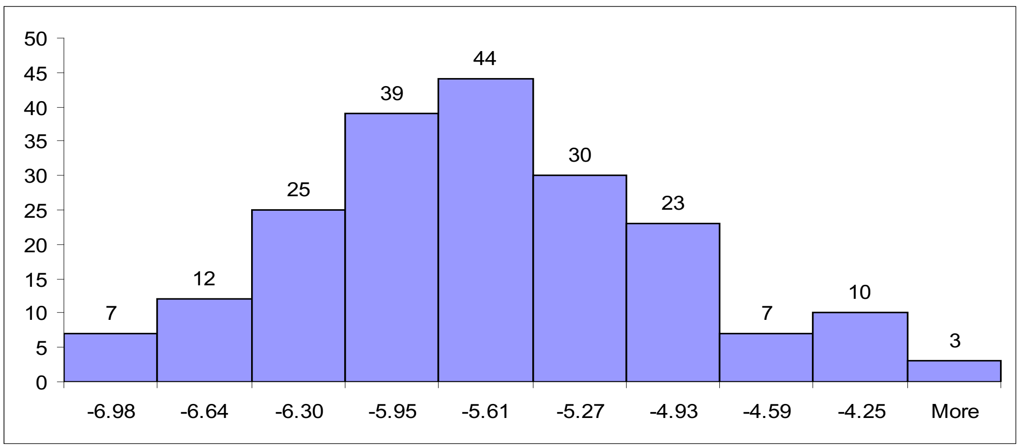

Fig. 1. Histogram of the quadratic bootstrap estimates $\hat{\beta}_0^*$ obtained for the resamples drawn from the original sample

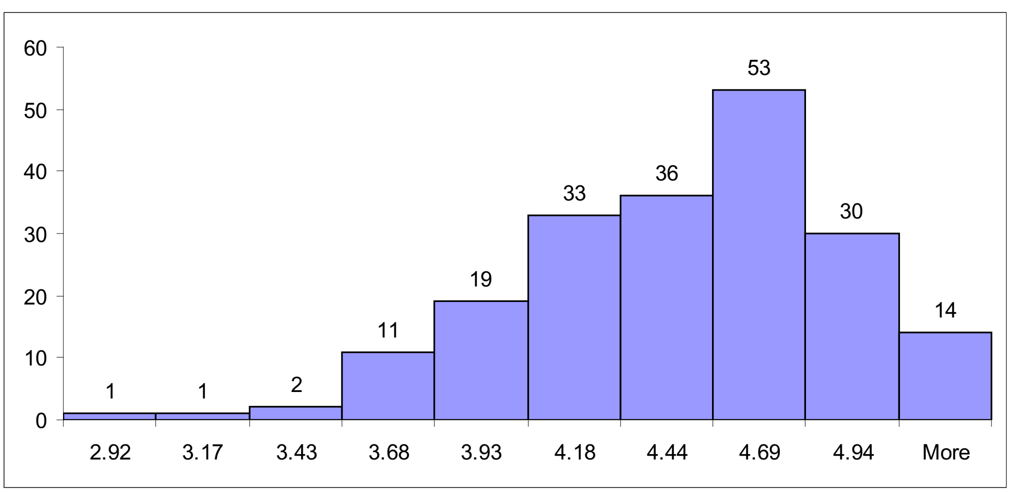

Fig. 2. Histogram of the quadratic bootstrap estimates $\hat{\beta}_1^*$ obtained for the resamples drawn from the original sample

The two models fitted based on ($\hat{\beta}_0$ = – 4.60, $\hat{\beta}_1$ = 3.46) and ($\hat{\beta}_0^*$ = – 5.78 and $\hat{\beta}_1^*$ = 4.36) are compared for their classifying abilities in terms of hold out data set. The classification is analyzed through a 2 × 2 classification table and the results are tabulated in Table 2. It has been found that both the models are having high classificatory power. However, the correct classification ratio, sensitivity and specificity for the logistic regression model fitted based on quadratic bootstrap estimates is more as compared to that of maximum likelihood estimates. Moreover, false positive rate and false negative rate are also less for quadratic bootstrap based fitted model as compare to the maximum likelihood based fitted model. Thus, quadratic bootstrap based estimates have been found to perform better marginally.

5. RESULTS OF THE SIMULATION STUDY

For studying further properties based on quadratic bootstrap estimation procedure, 2000 simulated samples each of size 108 have been generated from the following logistic regression model (values of the ML Estimates of the original sample have been chosen as the initial values for the model parameter $\hat{\beta}_0$ and $\hat{\beta}_1$ i.e. $\hat{\beta}_0$ = – 4.60 and $\hat{\beta}_1$ = 3.46). Thus the initial fitted logistic regression model is

$$\hat{\pi}_1 = P(\text{Y}=1) = \left(\frac{1}{1+e^{4.60-3.46X}}\right) \qquad (10)$$

where X is the single explanatory variable i.e. load given to the farm machinery which takes five different values as {0.00, 0.90, 1.80, 2.70, 3.60}; here, each level of X has been considered as a population. So, in total there are five populations. For the present study, the population sizes are 20, 24, 21, 19 and 24 respectively which together represent all the 108 observations. Now putting the value of X in Equation (10), probability that Y can take value one is obtained corresponding to each value of X Multiplying these probabilities by the corresponding number of observations under a particular population, the expected frequencies of occurrences of '1' in each of the populations can be calculated based on which we have generated 2000 simulated sample each of size 108. Thereafter, from each simulated data set, 200 bootstrap resamples have been taken and quadratic bootstrap estimates are constructed. Each of these 200 quadratic bootstrap estimators corresponding to a particular simulated sample have been used to obtain a bias corrected estimators $\left(\hat{\beta}_0^{bc}, \hat{\beta}_1^{bc}\right)$ as defined in Equation (5). Like this, for each of the remaining simulated samples, a bias corrected estimator has been constructed. In this manner, a total of 2000 such bias corrected estimates have been obtained. Along with bias corrected estimates $\left(\hat{\beta}_0^{bc}, \hat{\beta}_1^{bc}\right)$, MLE estimates $\left(\hat{\beta}_{0_{\text{MLE}}}, \hat{\beta}_{1_{\text{MLE}}}\right)$ have also been constructed for each of the corresponding simulated sample yielding 2000 MLE estimates. Thereafter, mean square error [defined Equation (8) and (9)] for both MLE and bias corrected estimators over all simulated samples have been calculated. For 2000 simulated samples, Table 3 shows simulated means, standard deviation and mean square error values of $\left(\hat{\beta}_{0_{\text{MLE}}}, \hat{\beta}_{1_{\text{MLE}}}\right)$ and bias corrected $\left(\hat{\beta}_0^{bc}, \hat{\beta}_1^{bc}\right)$ estimators. Perusal of Table 3 reveals that the bias correction decreases the variance, as the simulated standard deviation $\sigma\left(\hat{\beta}_0^{bc}\right)$ and $\sigma\left(\hat{\beta}_1^{bc}\right)$ for the bias corrected estimates are, smaller than the corresponding simulated values of standard deviation $\sigma\left(\hat{\beta}_{0_{\text{MLE}}}\right)$ and $\sigma\left(\hat{\beta}_{1_{\text{MLE}}}\right)$ for the MLE estimates respectively. Moreover, mean square errors (MSE) for bias corrected estimators $\left(\hat{\beta}_0^{bc}, \hat{\beta}_1^{bc}\right)$ have been reduced to a great extent as compared to the same of MLE estimates. Rather the efficiency factor of MLE $\left(\hat{\beta}_{0_{\text{MLE}}}, \hat{\beta}_{1_{\text{MLE}}}\right)$ is much less as compared to the bias corrected estimators $\left(\hat{\beta}_0^{bc}, \hat{\beta}_1^{bc}\right)$. The efficiency factor of $\hat{\beta}_{0_{\text{MLE}}}$ as compared to β_0^{bc} is 0.36 and the same for $\hat{\beta}_{1_{\text{MLE}}}$ as compared to is 0.31 respectively.

Table 3. Simulated means, standard deviation and mean square error values of maximum likelihood estimates ($\hat{\beta}_{0\,\mathrm{MLE}}$, $\hat{\beta}_{1\,\mathrm{MLE}}$) and bias corrected ($\hat{\beta}_0^{bc}$, $\hat{\beta}_1^{bc}$) estimates

$\hat{\beta}_0$ = - 4.60 (0.99) n = 108		$\hat{\beta}_1$ = 3.462 (0.71) n = 108	
$E(\hat{\beta}_{0\,\mathrm{MLE}})$	-5.77	$E(\hat{\beta}_{1\,\mathrm{MLE}})$	4.34
$E(\hat{\beta}_0^{bc})$	-4.17	$E(\hat{\beta}_1^{bc})$	3.11
$\sigma(\hat{\beta}_{0\,\mathrm{MLE}})$	1.02	$\sigma(\hat{\beta}_{1\,\mathrm{MLE}})$	0.79
$\sigma(\hat{\beta}_0^{bc})$	0.82	$\sigma(\hat{\beta}_1^{bc})$	0.56
MSE $(\hat{\beta}_{0\,\mathrm{MLE}})$	2.40	MSE $(\hat{\beta}_{1\,\mathrm{MLE}})$	1.41
MSE $(\hat{\beta}_0^{bc})$	0.86	MSE $(\hat{\beta}_1^{bc})$	0.43
$\frac{\mathrm{MSE}\,(\hat{\beta}_0^{bc})}{\mathrm{MSE}\,(\hat{\beta}_{0\,\mathrm{MLE}})}$	0.36	$\frac{\mathrm{MSE}\,(\hat{\beta}_1^{bc})}{\mathrm{MSE}\,(\hat{\beta}_{1\,\mathrm{MLE}})}$	0.31

6. DISCUSSION

Based on the present study, it can be said that, from estimation point of view, in terms of confidence intervals calculated for the estimates upon the original sample, quadratic bootstrap performs better than MLE as the length of confidence interval is smaller for the former as compare to the later. In terms of classificatory ability, although both the fitted model is having high classificatory ability, however, the model fitted based on quadratic bootstrap estimates performs marginally better than the fitted logistic model based on maximum likelihood estimates. Beside this, when simulation study is conducted, quadratic bootstrap based bias corrected estimates outperform MLE estimates with respect to mean square error and efficiency factor. The bias correction also decreases the variance. Thus the present study validate the theoretical results obtained in the literature based on a real experimental setup in the field of Agricultural Ergonomics.

In the context of the present study, it can be mentioned that, improved statistical models are always sought over and above the existing models to describe the aspects of the observed phenomena for better classification and prediction. Thus, while the existing logistic regression models may work well for certain situations, but may fail, in a slightly different set up, the improved ones which are bias corrected and robust can be used by subject matter specialists with more reliability. In this era of statistical computing both in terms of ease of speed and programming capabilities, such tasks can easily and readily be relegated to high speed computers and thus model fitting and analysis is no longer a problem nowadays. Rather applying improved logistic regression models as has been done in this study is worth the efforts because it gives a bias corrected estimator in addition to being more accurate.

ACKNOWLEDGEMENT

The authors are thankful to the Editor and the referees for the valuable comments on the previous draft of the manuscript. The facilities provided by the Director, ICAR-IASRI, New Delhi to carry out the research are duly acknowledged.

REFERENCES

Aerts, M. and Claeskens, G. (2001). Bootstrap tests for misspecified models, with application to clustered binary data. *Compu. Statist. Data Anal.,* **36,** 383-401.

Ayan, M.N.R. and Garcia, M.T.C. (2008). Prediction of University students' academic achievement by linear and logistic models. *The Spanish J. Psychol.,* **11(1)**, 275-288.

Bhowmik, A., Ramasubramanian, V., Chandrahas and Kumar, A. (2011). Logistic regression for classification in agricultural ergonomics. *Adv. Appl. Res.,* **3(2)**, 163-170.

Claeskens, G. and Aerts, M. (2000). Bootstrapping local polynomial estimators in likelihood-based models. *J. Statist. Plann. Inf.,* **86**, 63-80.

Claeskens, G., Aerts, M. and Molenberghs, G. (2003). A quadratic bootstrap method and improved estimation in logistic regression. *Statist. Probab. Lett.,* **61**, 383-394.

Gent, D.H., Schwartz, H.F., Ishimaru, C.A., Louws, F.J., Cramer, R.A. and Lawrence, C.B. (2003). Polyphasic characterization of *Xanthomonas* strains from onion. *Phytopathology***, 94(2)**, 184-195.

Johnson, D.A., Alldredge, R. and Vakoch, D.L. (1996). Potato late blight forecasting models for the semiarid environment of South-Central Washington. *Phytopathology,* **86**, 480-484.

Lee, K.W. (1990). Bootstrapping logistic regression models with random regressors. *Comm. Statist. - Theory Methods,* **19**, 2527–2539.

Mila, A.L., Carriquiry, A.L. and Yang, X.B. (2004). Logistic regression modeling of prevalence of soybean sclerotinia stems rot in the North-Central region of the United States. *Phytopathology,* **94**, 102-110.

Misra, A.L., Prakash, O. and Ramasubramanian V. (2002). Forewarning powdery mildew caused to *Odium mangiferae* in mango (*Mangifera indica*) using logistic regression models. *Ind. J. Agric. Sci.,* **74(2)**, 84-87.

Ram, R., Kumar, A., Singh, A.K., Jha, S.K. and Ramasubramanian V. (2008).Ergonomic evaluation of foot operated rotary power generation. *J. Agric. Engg.,* **45(3)**, 12-18.

Swapnepoel, C.J. and Frangos, C.C. (1994). Bootstrap confidence intervals for slope parameter of a logistic model. *Comm. Statist.: Sim. Comput.* **23(4)**, 1115-26.

Tsien, C.L., Fraser, H.S.F., Long, W.J., Kennedy, R.L. (1998). Using classification tree and logistic regression methods to diagnose myocardial infarction. *Medinfo,* **98**, 493-497.

Vergara, M. and Page, A. (2002). Relationship between comfort and back posture and mobility in sitting-posture. *Appl. Ergon.,* **33(1)**, 1-8.