11-2014

Comparison of Estimators in GLM with Binary Data

D. M. Sakate
Shivaji University, Kolhapur, India, dms.stats@gmail.com

D. N. Kashid
Shivaji University, Kolhapur, Maharashtra, India., dnk_stats@unishivaji.ac.in

Comparison of Estimators in GLM with Binary Data

D. M. Sakate
Shivaji University
Kolhapur, India

D. N. Kashid
Shivaji University
Kolhapur, India

Maximum likelihood estimates (MLE) of regression parameters in the generalized linear models (GLM) are biased and their bias is non negligible when sample size is small. This study focuses on the GLM with binary data with multiple observations on response for each predictor value when sample size is small. The performance of the estimation methods in Cordeiro and McCullagh (1991), Firth (1993) and Pardo et al. (2005) are compared for GLM with binary data using an extensive Monte Carlo simulation study. Performance of these methods for three real data sets is also compared.

Keywords: Binomial regression, modified score function, bias corrected MLE, Minimum ϕ-divergence estimation, Monte Carlo Simulation

Introduction

Generalized linear models (GLM) are frequently used to model small to medium size data. In case of binomial distributed response, logistic regression finds application to model the relationship between response and predictors. Maximum likelihood estimation (MLE) is usually used to fit a logistic regression model. It is well known that under certain regularity conditions, MLE of regression coefficients are consistent and asymptotically normal. However, for finite sample sizes, MLE tend to overestimate with an absolute bias that tends to increase with the magnitude of the parameter and with the ratio of the number of parameters to the number of observations. The bias in MLE decreases with the sample size and goes to zero as sample size tends to infinity. See Byth and McLachlan, (1978), Anderson and Richardson (1979), McLachlan (1980), Pike et al. (1980), Breslow (1981) and Hauck (1984) for the details. As a consequence, methods taking care of bias were explored. Jackknifed MLE and its versions and methods based on approximation

of bias using Taylor series expansion are widely studied in the literature. See Bull et al. (1994) and references therein. Cordeiro and McCullagh (1991) proposed second order unbiased MLE in GLM. Further, to simultaneously tackle the problem of bias and separation, Firth (1993) modified the score function to estimate the parameters unbiasedly up to first order. Maiti and Pradhan (2008) empirically established the superiority of these two methods over conditional maximum likelihood estimator in non-separable case through extensive simulation study.

In the last decade, the minimum distance estimators have gained importance in many fields of statistics. Read and Cressie (1988) and Pardo (2006) outlined the use and importance of the ϕ-divergence measures in statistics. Pardo et al. (2005) proposed the minimum ϕ-divergence estimator or minimum distance estimator based on the family of power divergence (Cressie and Read, 1984) characterized by a tuning parameter λ for estimation of regression coefficients in logistic regression. The minimum ϕ-divergence estimator is a generalization of MLE ($\lambda = 0$). Other distance estimators like minimum chi-square estimator ($\lambda = 1$) and minimum Hellinger distance estimator ($\lambda = -1/2$) are particular cases as well. An extensive simulation study in Pardo et al. (2005) and Pardo and Pardo (2008) to choose among the estimators in logistic regression concluded that 2/3 is a good choice for λ. Hence, minimum ϕ-divergence estimator with $\lambda = 2/3$ emerged as an alternative to MLE in the sense of MSE for small size. The comparison of the minimum distance estimators with those taking care of bias remains the untouched problem of interest.

Estimation in logistic regression

Let Z be a response binary random variable taking value 1 or 0, generally referred to as "success" or "failure" respectively. Let k explanatory variables $x \in \mathbb{R}^k$ are observed along with the response variable. $\pi(x) = P(Z = 1 \mid x \in \mathbb{R}^k)$ represents the conditional probability, of the value 1 given $x \in \mathbb{R}^k$. Let X be the $N \times (k+1)$ matrix with rows $x_i = (x_{i0}, x_{i1}, \ldots, x_{ik})$, $i = 1, \ldots, N$ where $x_{i0} = 1$, $\forall\ i$. The logistic regression model is defined by the conditional probability

$$\pi(x_i) = \frac{\exp\left\{\beta_0 + \sum_{j=1}^{k}\beta_j x_{ij}\right\}}{1 + \exp\left\{\beta_0 + \sum_{j=1}^{k}\beta_j x_{ij}\right\}}. \tag{1}$$

For more discussion on logistic regression see Hosmer and Lemeshow (1989) and Agresti (1990).

In laboratory or controlled setting, many individuals share same values for their explanatory variables. In other words, for each value of the explanatory variables there are several observed values of the random variable Z. The focus here is on this situation. The notations described earlier are required to be changed slightly. For this, the notations used in Pardo et al. (2005) were followed. Let there be I distinct values of $x_i = (x_{i0}, x_{i1}, \ldots, x_{ik})$, $i = 1, 2, \ldots, I$. It is assumed that, for each x_i, there is a binomial random variable $Y_i \equiv \sum_{i=1}^{n_i} Z_i$ with parameters n_i and $\pi(x_i)$. The values $n_{i1}, \ldots, n_{I1}$ are the observed values of the random variables $Y_1, \ldots, Y_I$, representing the number of successes in $n_1, \ldots, n_I$ trials respectively when the explanatory variables are fixed. This divides the entire sample of size N into I subgroups each of size n_i so that $= \sum_{i=1}^{I} n_i$. Because Z_i's are independent, Y_i's are also independent. Thus, the likelihood function for the logistic regression model is given by

$$L(\beta_0,\ldots,\beta_k) = \prod_{i=1}^{I} \binom{n_i}{n_{i1}} \pi\left(x_i^T \beta\right)^{n_{i1}} \left(1 - \pi\left(x_i^T \beta\right)\right)^{n_i - n_{i1}} \tag{2}$$

The MLE, $\hat{\beta}$ is derived as a solution to score equation

$$l'(\beta) = U(\beta) = \mathbf{0} \tag{3}$$

where $l(\beta) = \log L(\beta)$ is the log likelihood function.

Second order bias corrected MLE

As discussed earlier, there are various methods which give rise to bias corrected versions of MLE in logistic regression (Anderson and Richardson, 1979; McLachlan, 1980; Schaefer, 1983; Copas, 1988 and Cordeiro and McCullagh, 1991). Cordeiro and McCullagh (1991) derived an expression for the first order bias using Taylor series expansion. Let μ be the mean of response variable. The n^{-1} bias of $\hat{\beta}$ in GLM is given by

$$B_1(\beta) = \left(X^T W X\right)^{-1} X^T \xi$$

where $W = diag\{V^{-1}\boldsymbol{\mu}'^2\}$ is an $n \times n$ weight matrix and

$$\boldsymbol{\xi} = -(2\psi)^{-1}W^{-1}H_d\boldsymbol{F}\mathbf{1}.$$

ψ is the dispersion parameter of the GLM, H_d is an $n \times n$ diagonal matrix with the elements of $H = X(X^TWX)^{-1}X^T$ and $F = diag\{V^{-1}\boldsymbol{\mu}'\boldsymbol{\mu}''\}$ is also an $n \times n$ diagonal matrix. The MLE of $B_1(\boldsymbol{\beta})$ is then subtracted to obtain the second order bias corrected estimate

$$\hat{\boldsymbol{\beta}}^{(2)} = \hat{\boldsymbol{\beta}} - B_1(\hat{\boldsymbol{\beta}}).$$

Modified score function method

The bias in MLE is due to unbiasedness and curvature of score function. The score function is linear in case of normal error regression and hence MLE is unbiased. Firth (1993) modified the score function by introducing small bias in score function. This produces a separation resistant estimator with zero first order bias.

The modified score function is defined as

$$U^*(\boldsymbol{\beta}) = U(\boldsymbol{\beta}) - i(\boldsymbol{\beta})B_1(\boldsymbol{\beta}) = 0$$

where $i(\boldsymbol{\beta})$ is the Fisher information matrix.

The solution point of the modified score equation locates a stationary point of

$$l^*(\boldsymbol{\beta}) = l(\boldsymbol{\beta}) + \tfrac{1}{2}\log|i(\boldsymbol{\beta})|$$

or equivalently, of the penalized likelihood function

$$L^*(\boldsymbol{\beta}) = L(\boldsymbol{\beta}) + |i(\boldsymbol{\beta})|^{1/2}$$

where $|i(\boldsymbol{\beta})|^{1/2}$ is the Jeffreys (1946) invariant Prior for the problem.

In GLM with Binary data, the modified score function is (Firth, 1993)

$$U^*(\boldsymbol{\beta}) = U(\boldsymbol{\beta}) - X^TW\boldsymbol{\xi}$$

In logistic regression, $W\xi$ has i^{th} element h_i $(\pi_i - 1/2)$ and h_i is the ith diagonal element of the hat matrix. The solution $U^*(\beta) = 0, \hat{\beta}^{MS}$ is free from the first order bias.

Minimum ϕ-Divergence Estimation

Let $\pi_{i1} = \pi(\mathbf{x}_i^T \beta)$ and $\pi_{i2} = 1 - \pi(\mathbf{x}_i^T \beta), n_{i2} = n_i - n_{i1}$ and $N = \sum_{i=1}^{I} n_i$. To maximize (2) is equivalent to minimizing the Kullback divergence measure between the probability vectors

$$\hat{p} = \left(\hat{p}_{11}, \hat{p}_{12}, \ldots, \hat{p}_{I1}, \hat{p}_{I2} \right)^T = \left(\frac{n_{11}}{N}, \frac{n_{12}}{N}, \ldots, \frac{n_{I1}}{N}, \frac{n_{I2}}{N} \right)^T \text{ and }$$

$$p(\beta) = \left(p_{11}(\beta), p_{12}(\beta), \ldots, p_{I1}(\beta), p_{I2}(\beta) \right)^T = \left(\pi_{11} \frac{n_1}{N}, \pi_{12} \frac{n_1}{N}, \ldots, \pi_{I1} \frac{n_I}{N}, \pi_{I2} \frac{n_I}{N} \right)^T.$$

MLE for the GLM parameter β can be defined by

$$\hat{\beta} = \arg\min_{\beta \in \Theta} D_{Kullback} \left(\hat{p}, p(\beta) \right), \tag{4}$$

where the Kullback divergence measure is given by (see Kullback, 1985)

$$D_{Kullback} \left(\hat{p}, p(\beta) \right) = \sum_{j=1}^{2} \sum_{i=1}^{I} \hat{p}_{ij} \log \left(\frac{\hat{p}_{ij}}{p_{ij}(\beta)} \right).$$

This measure is a particular case of the ϕ-divergence defined by Csiszar (1963) and Ali and Silvey (1966),

$$D_\phi \left(\hat{p}, p(\beta) \right) = \sum_{j=1}^{2} \sum_{i=1}^{I} p_{ij}(\beta) \phi \left(\frac{\hat{p}_{ij}}{p_{ij}(\beta)} \right); \phi \in \Phi, \tag{5}$$

where Φ is the class of all convex functions $\phi(x), x > 0$ such that at $x = 1, \phi(1) = \phi'(1) = 0, \phi''(1) > 0$ and at $x = 0, 0 \; \phi(0/0) = 0$ and $0 \; \phi(p/0) = p \lim_{u \to \infty} \phi(u)/u$. For more details, see Vajda (1989) and Pardo (2006).

Cressie and Read (1984) introduced an important family of ϕ-divergences called the power divergence family

$$\phi_\lambda(x) = (\lambda(\lambda+1))^{-1}(x^{\lambda+1} - x); \lambda \neq 0, \lambda \neq -1,$$
$$\phi_0(x) = \lim_{\lambda \to 0} \phi_\lambda(x) = x\log(x) - x + 1, \tag{6}$$
$$\phi_{-1}(x) = \lim_{\lambda \to -1} \phi_\lambda(x) = -\log(x) - x - 1.$$

It is interesting to note that

$$D_{\phi_0}(\hat{p}, p(\beta)) = D_{Kullback}(\hat{p}, p(\beta)). \tag{7}$$

The minimum ϕ-divergence estimator (Pardo et al., 2005) in logistic regression is given by

$$\hat{\beta}_\phi = \arg\min_{\beta \in \Theta} D_\phi(\hat{p}, p(\beta)) \tag{8}$$

Methodology

Performance comparison for real data

Usefulness of any method can be established only when it can be applied to a real data. To this end, the performance of these methods is compared for prediction purposes when real data is used. Three real data sets are employed as examples. The numbers of predictors in each real data are not same. A single predictor, two predictor and multiple predictor situations are considered in real data to compare the prediction performance of the methods.

Example 1: Single Predictor Case

First, for the single predictor case, pneumoconiosis data (Montgomery et. al., 2006; pp. 449) concerning the proportion of coal miners who exhibit symptoms of severe pneumoconiosis and the numbers of years of exposure (X_1) is analyzed. The data includes $n = 8$ observations on number of years of exposure, number of severe cases and total number of miners. Table 1 presents the estimated regression coefficients using the methods discussed earlier.

Table 1. Estimates of regression coefficients: Example 1

Predictors	$\hat{\beta}$	$\hat{\beta}^{(2)}$	$\hat{\beta}_{\phi}$	$\hat{\beta}^{MS}$
Intercept	-4.55621	-4.47047	-4.54838	-4.48677
# Years of Exposure	0.07889	0.07707	0.08753	0.07747

Example 2: Two Predictors Case

The two predictors case example uses ingots data, which appeared in Cox and Snell (1989, pp. 10-11). This data is a result of an industrial experiment concerning steel ingots. It consists of 19 observations on the number of ingots not ready for rolling out of certain number of trials tested for a number of heating time (X_1) and soaking time (X_2). This data is also analyzed by Pardo et al. (2005) to illustrate the use of minimum ϕ-divergence estimator. Table 2 presents the estimates of regression coefficients in the binomial logistic regression.

Table 2. Estimates of regression coefficients: Example 2

Predictors	$\hat{\beta}$	$\hat{\beta}^{(2)}$	$\hat{\beta}_{\phi}$	$\hat{\beta}^{MS}$
Intercept	-5.51316	-5.42360	-4.88651	-5.47685
Heating time	0.07688	0.07573	0.06881	0.07690
Soaking time	0.07201	0.11485	0.04469	0.10876

Example 3: Multiple Predictors Case

In this next example, a real data with more than two predictors is analyzed and considers data (Andersen 1997, pp. 171) used by Pardo and Pardo (2008) to illustrate the variable selection method based on minimum ϕ-divergence estimator. The data consists of observations on 6 objective indicators (X_1, …, X_6) of the actual indoor climate in 10 classrooms of a Danish Institute, the number of students in the class and the number of yes-answers to the question whether they felt that the indoor climate at the moment was pleasant or not so pleasant. Table 3 presents the estimates of regression coefficients in the binomial logistic regression. The minimum ϕ-divergence estimate of coefficient of X_3 and X_5 differ in magnitude as compared to estimates from other three methods to a larger extent.

Table 3. Estimates of regression coefficients: Example 3

Predictors	$\hat{\beta}$	$\hat{\beta}^{(2)}$	$\hat{\beta}_\phi$	$\hat{\beta}^{MS}$
Intercept	5.75029	5.48737	7.15570	5.50296
X_1	0.53530	0.51090	1.29870	0.51380
X_2	-0.51320	-0.49247	-1.15230	-0.49442
X_3	9.04758	8.65945	19.28910	8.69477
X_4	0.64191	0.61565	1.42370	0.61780
X_5	8.93732	8.67950	25.20410	8.67973
X_6	-0.04478	-0.04263	-0.07700	-0.04283

It is not possible decide between the estimators merely by looking at the estimated regression coefficients. To compare the performance in each example, the predicted probabilities of success using each estimator were computed. As a measure of discrepancy between estimated and observed probability of success, the Pearson chi-square defined as

$$\text{Pearson chi-square} = \sum_{i=1}^{I}\left(\frac{P_i - \hat{\pi}_{i1}^c}{\sqrt{n_i\hat{\pi}_{i1}^c\hat{\pi}_{i2}^c}}\right)^2$$

was used, where, $P_i = \frac{n_{i1}}{n_i}$ and c represents the method used to estimate probability of success. The Pearson chi-square corresponding to estimators considered for all the examples are listed in the Table 4. The Pearson chi-square for $\hat{\beta}_\phi$ is smallest in case of Examples 1 and 3. For Example 2, value of Pearson chi-square is smallest for $\hat{\beta}$. It reveals that the performance of minimum ϕ-divergence estimators to predict probabilities of success in binomial logistic regression applied to real data is better than MLE and its bias corrected versions.

Table 4. Pearson chi-square

Estimator	$\hat{\beta}$	$\hat{\beta}^{(2)}$	$\hat{\beta}_{\phi}$	$\hat{\beta}^{MS}$
Example 1	0.0058	0.0057	0.0056	**0.0019**
Example 2	**0.7570**	0.8733	0.8648	0.8682
Example 3	0.2419	0.2393	0.2394	**0.0216**

Monte-Carlo Simulation Study

Estimation methods were compared using Monte-Carlo simulation; a two predictor binomial logistic regression model including an intercept was considered. The design matrix is of order 11×3 with first column as ones. The other two columns contain random numbers from two independent standard uniform distributions. To generate observations on response variable, two different parameter structures were considered and accordingly two different models were defined as Model I: $\beta = (1, 2, -3)$; Model II: $\beta = (1, -1.5, 2)$.

Table 5. AMSE with its SD

		$\hat{\beta}$	$\hat{\beta}^{(2)}$	$\hat{\beta}_{\phi}$	$\hat{\beta}^{MS}$
	n^1	2.4014 (0.5287)	2.4195 (0.5235)	2.4173 (0.5244)	0.2902 (0.3909)
	n^2	2.4055 (0.6030)	2.4355 (0.5951)	2.4298 (0.5956)	0.4202 (0.5254)
	n^3	2.3944 (0.4684)	2.4068 (0.4646)	2.4055 (0.4654)	0.1768 (0.2292)
	n^4	2.4228 (0.5837)	2.4553 (0.5739)	2.4512 (0.5759)	0.4210 (0.4849)
	n^5	2.4455 (0.7145)	2.4945 (0.6952)	2.4873 (0.6992)	0.7497 (1.0160)
Model I	n^6	2.4371 (0.7172)	2.4811 (0.6996)	2.4757 (0.7032)	0.7200 (0.9379)
	n^7	2.4201 (0.5828)	2.4518 (0.5755)	2.4464 (0.5761)	0.4432 (0.5208)
	n^8	2.4332 (0.5633)	2.4562 (0.5566)	2.4537 (0.5580)	0.3410 (0.3640)
	n^9	2.5535 (0.9460)	2.6306 (0.8986)	2.6189 (0.9091)	1.5971 (5.8009)
	n^{10}	2.6685 (1.1593)	2.7568 (1.0321)	2.7364 (1.0605)	3.1885 (6.3671)

Table 5, continued

		$\hat{\beta}$	$\hat{\beta}^{(2)}$	$\hat{\beta}_{\phi}$	$\hat{\beta}^{MS}$
	n^1	1.8489	1.8528	1.8524	0.3023
		(0.2223)	(0.2206)	(0.2207)	(0.3825)
	n^2	1.8673	1.8732	1.8723	0.5364
		(0.2760)	(0.2722)	(0.2725)	(0.7490)
	n^3	1.8511	1.8537	1.8535	0.2053
		(0.1868)	(0.1857)	(0.1858)	(0.3301)
	n^4	1.8588	1.8653	1.8648	0.5813
		(0.2877)	(0.2828)	(0.2833)	(0.8713)
	n^5	1.8819	1.8913	1.8903	0.8732
Model II		(0.3377)	(0.3297)	(0.3305)	(1.1947)
	n^6	1.8753	1.8840	1.8833	0.7925
		(0.3233)	(0.3165)	(0.3172)	(0.9626)
	n^7	1.8644	1.8705	1.8695	0.5432
		(0.2984)	(0.2935)	(0.2942)	(0.6855)
	n^8	1.8500	1.8550	1.8547	0.4296
		(0.2526)	(0.2498)	(0.2501)	(0.8013)
	n^9	1.9098	1.9237	1.9221	1.8013
		(0.4462)	(0.4260)	(0.4284)	(3.5744)
	n^{10}	1.9155	1.9282	1.9268	2.1006
		(0.4511)	(0.4276)	(0.4305)	(7.0825)

Table 6. Average absolute bias

		$\hat{\beta}$	$\hat{\beta}^{(2)}$	$\hat{\beta}_{\phi}$	$\hat{\beta}^{MS}$
	n^1	0.2556	0.2534	0.2535	0.0023
	n^2	0.2580	0.2539	0.2545	0.0048
	n^3	0.2546	0.2531	0.2531	0.0018
	n^4	0.2550	0.2508	0.2511	0.0002
Model I	n^5	0.2576	0.2508	0.2513	0.0065
	n^6	0.2640	0.2575	0.2580	0.0108
	n^7	0.2586	0.2543	0.2548	0.0055
	n^8	0.2516	0.2487	0.2488	0.0094
	n^9	0.2606	0.2471	0.2485	0.0210
	n^{10}	0.2721	0.2475	0.2510	0.0010
	n^1	0.5785	0.5780	0.5780	0.0075
	n^2	0.5793	0.5785	0.5786	0.0149
	n^3	0.5780	0.5777	0.5777	0.0031
	n^4	0.5788	0.5779	0.5779	0.0159
Model II	n^5	0.5792	0.5778	0.5779	0.0258
	n^6	0.5787	0.5774	0.5774	0.0120
	n^7	0.5782	0.5773	0.5774	0.0093
	n^8	0.5783	0.5776	0.5776	0.0079
	n^9	0.5842	0.5813	0.5814	0.0309
	n^{10}	0.5810	0.5781	0.5782	0.0182

Consider the following 10 different combinations of number of trials

$$n^1 : 15,15,15,15,30,30,30,30,40,40,40$$
$$n^2 : 5,5,5,5,15,15,15,15,40,40,40$$
$$n^3 : 40,40,40,40,40,40,40,40,40,40,40$$
$$n^4 : 10,10,10,10,20,20,20,20,15,15,15$$
$$n^5 : 10,10,10,10,5,5,5,5,15,15,15$$
$$n^6 : 10,10,10,10,10,10,10,10,10,10,10$$
$$n^7 : 5,5,5,5,30,30,30,30,15,15,15,15$$
$$n^8 : 20,20,20,20,20,20,20,20,20,20,20$$
$$n^9 : 5,5,5,5,5,5,5,5,5,5,5$$
$$n^{10} : 5,5,5,5,5,5,5,5,4,4,4$$

The observations on response variable are random numbers from $B(n_i, \pi_{i1})$. In this way, 20 models were generated differing in parameter structure and structure of number of trials. Unknown regression coefficients were estimated using four methods including MLE. To compute the minimum ϕ-divergence estimate, the power divergence family in (6) with $\lambda = 2/3$ were used as suggested in Pardo et. al. (2005). Each model was simulated 1,000 times and average MSE (AMSE) and average absolute bias in estimate due to each estimation method are reported in the Tables 5 and 6. The figures in parentheses represent standard deviation (SD) of MSE. The AMSE and average absolute bias were computed using the following formulae

$$\text{AMSE} = \frac{1}{1000} \sum_{i=1}^{1000} \frac{1}{3} \sum_{j=0}^{2} \left(\hat{\beta}_{ij}^c - \beta_j \right)^2.$$

$$\text{Average absolute bias} = \frac{1}{1000} \sum_{i=1}^{1000} \frac{1}{3} \sum_{j=0}^{2} \left| \hat{\beta}_{ij}^c - \beta_j \right|.$$

Results

It is evident from the Tables 5 and 6, the minimum ϕ-divergence estimator has smaller MSE and bias as compared to others for all combinations of number of trials. For a small magnitude of number of trials, as in case of last combination, the

AMSE of all the estimators is more or less same; however, variability in the minimum ϕ-divergence estimate is quite high. The estimate based on modified score function and second order bias corrected MLE are close enough to uncorrected MLE in this setting. The bias correction obtained as such is negligible. From Tables 5 and 6, performance of minimum ϕ-divergence estimator is better than the others for all but last two combinations of number of trials i.e., n^9 and n^{10}.

The variation in MSE and absolute bias averaged over the three regression coefficients is shown in Figures 1 and 2 respectively for n^1 and n^{10}. Although the motivation behind defining the minimum ϕ-divergence estimator was altogether different, it performs better than the bias corrected versions of MLE. This makes the minimum ϕ-divergence estimator an attractive alternative to MLE as well as is its bias corrected versions in binomial logistic regression.
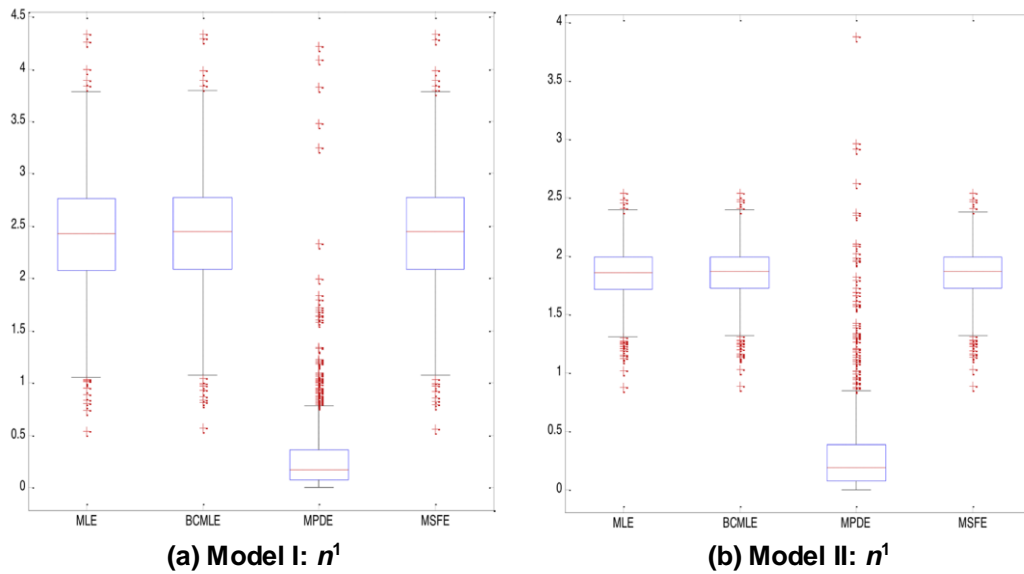

(a) Model I: n^1   **(b) Model II: n^1**

Figure 1 (a, b). Box Plot of MSE of estimates averaged over three regression coefficients.

(c) Model I: n^{10}   **(d) Model II: n^{10}**
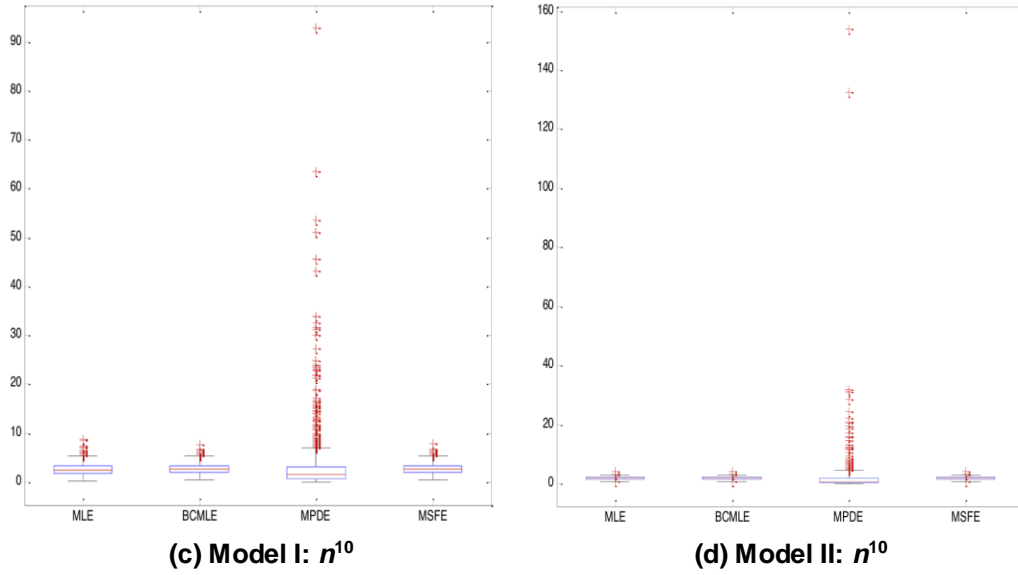
Figure 1 (c, d). Box Plot of MSE of estimates averaged over three regression coefficients.

(a) Model I: n^1

(b) Model I: n^{10}
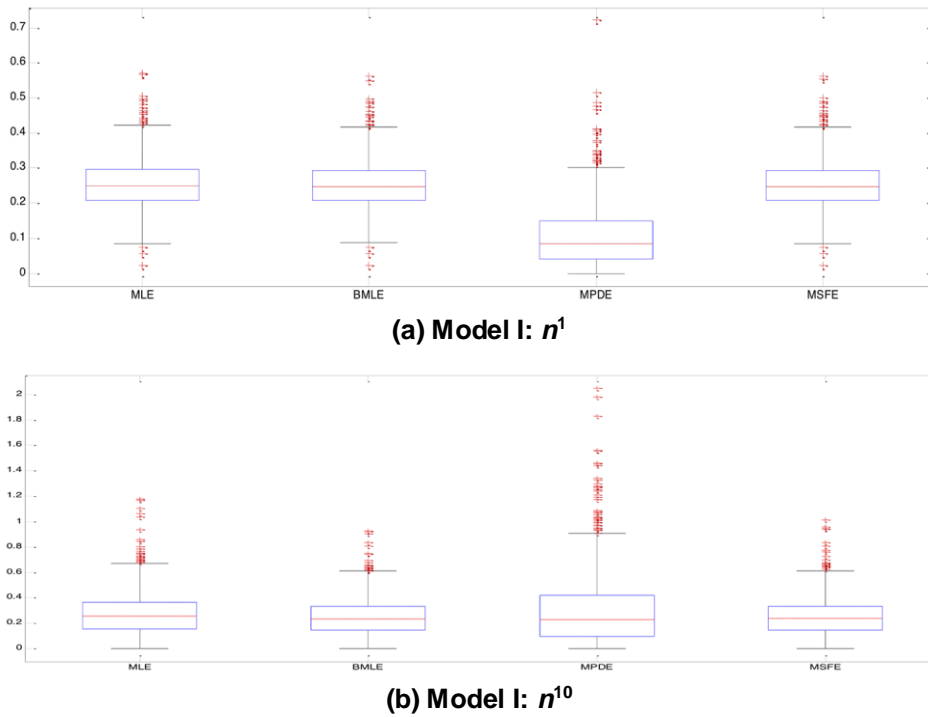
Figure 2 (a, b). Box Plot of bias of estimates averaged over three regression coefficients.

(c) Model II: n^1

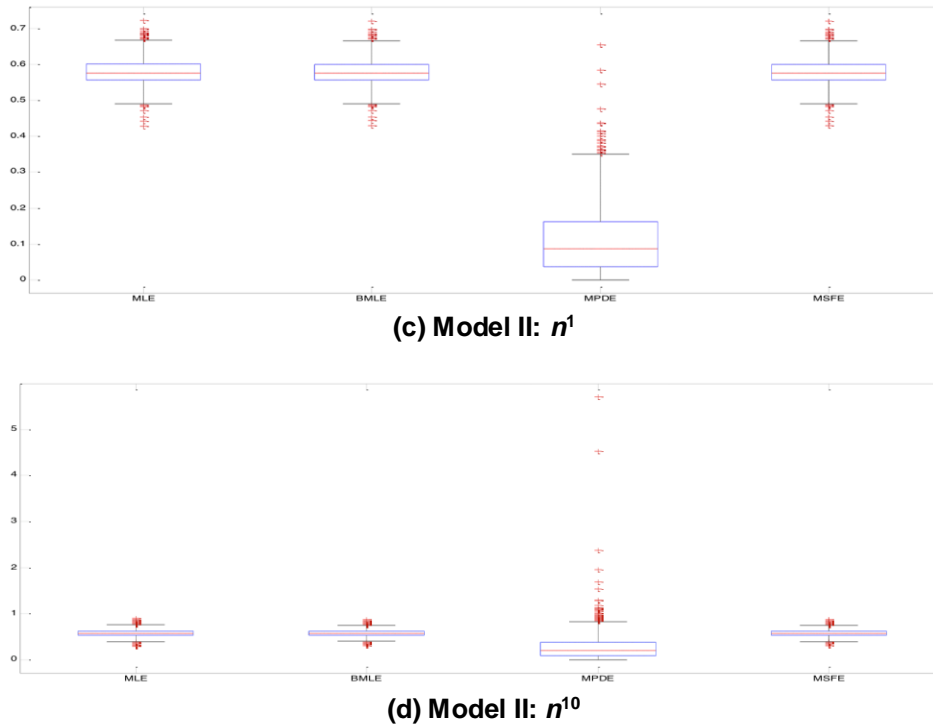**(d) Model II: n^{10}**

Figure 2 (c, d). Box Plot of bias of estimates averaged over three regression coefficients.

Conclusion

The performance of some of the estimators belonging to two different classes, i.e., minimum distance estimators and bias corrected MLE in a binomial logistic regression model, was compared. Three real data examples from different fields followed by a Monte Carlo simulation study were used to illustrate the comparisons. Results show that second order bias corrected MLE and estimates obtained using modified score function method lead to an estimate, which is same as MLE when number of trials is large. From this comparison study it may be concluded that, for a number of trials greater than 5, minimum ϕ-divergence estimator is an attractive alternative to MLE as well as bias corrected and modified score function method.

References

Agresti, A. (1990). *Categorical data analysis*. New York: John Wiley and Sons.

Ali, S. M. & Silvey, S. D. (1966). A general class of coefficients of divergence of one distribution from another. *Journal of the Royal Statistical Society, Series B, 26*, 131-142.

Andersen, E. B. (1997). *Introduction to the statistical analysis of categorical data I*. New York: Springer.

Anderson, J. A. & Richardson, S. C. (1979). Logistic discrimination and bias correction in maximum likelihood estimation. *Technometrics, 21*, 71-78.

Breslow, N. (1981). Odds ratio estimators when the data are sparse. *Biometrika, 68*, 73-84.

Bull, S. S., Hauck, W. W. & Greenwood, C. M. T. (1994). Two-step Jackknife bias reduction for logistic regression MLEs. *Communication in Statistics – Simulation and Computation, 23*, 59-88.

Byth, K. & McLachlan, G. J. (1978). The biases associated with maximum likelihood methods of estimation of the multivariate logistic risk function. *Communication in Statistics – Theory and Methods, A7*, 877-890.

Copas, J. B. (1988). Binary regression models for contaminated data with discussion. *Journal of the Royal Statistical Society, Series B, 50*, 225-265.

Cordeiro, G. M. & McCullagh, P. (1991). Bias correction in generalized linear models. *Journal of the Royal Statistical Society, Series B, 53*, 629-643.

Cox, D. R. & Snell, E. J. (1989). *Analysis of Binary Data*. London: Chapman and Hall.

Cressie, N. A. C. & Read, T. R. C. (1984). Multinomial goodness of fit tests. *Journal of the Royal Statistical Society, Series B, 46*, 440-464.

Csiszár, I. (1963). Eine Informationtheorestiche Ungleichung und ihre Anwendung anf den Beweis der Ergodizität Markoffshen Ketten. *Publications of the mathematical Institute of Hungarian Academy of Sciences, Series A, 8*, 84–108.

Firth, D. (1993) Bias Reduction of Maximum Likelihood Estimates. *Biometrika, 80*, 27-38.

Hauck, W. W. (1984). A comparative study of conditional maximum likelihood estimation of a common odds ratio. *Biometrics, 40*, 1117-1123.

Hosmer, D. W. & Lemeshow, S. (1989). *Applied Logistic Regression*. New York: John Wiley and Sons.

Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London 186*, 453–461.

Kullback, S. (1985). 'Kullback information.' In S. Kotz & N. L. Johnson (Eds.). *Encyclopedia of Statistical Sciences, 4*, 421-425. New York: John Wiley and Sons.

Maiti, T. & Pradhan, V. (2008). A comparative study of the bias corrected estimates in logistic regression. *Statistical Methods in Medical Research, 17*, 621-634.

McCullagh P. & Nelder J. A. (1989). *Generalized Linear models* (Second ed.). London: Chapman and Hall.

McLachlan, G. J. (1980). A note on bias correction in maximum likelihood estimation with logistic discrimination. *Technometrics, 22*, 621-627.

Montgomery, D. C., Peck, E. A & Vining, G. G. (2006). *Introduction to linear regression analysis*. New York: John Wiley and Sons.

Pardo, J. A., Pardo, L., & Pardo, M. C. (2005). Minimum ϕ-divergence estimator in logistic regression models. *Statistical Papers, 47*, 91-108.

Pardo, J. A. & Pardo, M. C. (2008). Minimum ϕ-divergence estimator and ϕ-divergence statistics in generalized linear models with binary data. *Methodology and Computing in Applied Probability, 10*, 357-379.

Pardo, L. (2006). *Statistical inference based on divergence measures*. New York: Taylor and Francis Group, LLC.

Pike, M. C., Hill, A. P., & Smith, P. G. (1980). Bias and efficiency in logistic analysis of stratified case control studies. *International Journal of Epidemiology, 9*, 705-724.

Read, T. R. C. & Cresie, N. (1988). *Goodness of fit statistics for discrete multivariate data*. New York: Springer.

Schaefer, R. L. (1983). Bias correction in maximum likelihood logistic regression, *Statistics in Medicine, 2*, 71-78.

Vajda, I. (1989). *Theory of statistical inference and information*. Boston: Kluwer.