ISSN 1684-8403
Journal of Statistics
Volume 16, 2009, pp. 1-11

Consistency of the Maximum Likelihood Estimator in Logistic Regression Model: A Different Approach

Mamunur Rashid[1] and Naima Shifa[2]

Abstract

This article investigates the consistency of maximum likelihood estimators for the logistic regression model using a different approach. In this approach, we verify all the regularity conditions for the logistic regression model needed for consistency of the maximum likelihood estimators. The performance of the maximum likelihood estimators for consistency in the logistic regression model is also examined via standard Monte Carlo simulation study.

Keywords

Logistic regression, Consistency, Maximum likelihood estimator, Monte Carlo simulation

1. Introduction

A common name for a regression model for binary response variables is the logistic regression model, which has been widely used in the physical, biomedical, and behavioral sciences (Mehta et al., 2000). Let Y be a binary variable and let X be the associated $p \times 1$ vector of explanatory variables. Then the standard logistic regression model assumes the following model:

$$P\,(Y = 1 \mid X = x) = \frac{e^{\beta_0 + \beta^T x}}{1 + e^{\beta_0 + \beta^T x}}$$

[1] Department of Mathematics and Statistics, Ohio Northern University, Ada, OH 45810, USA
E-mail: m-rashid@onu.edu
[2] Department of Mathematics, DePauw University, Greencastle, IN 46135, USA
E-mail: naimashifa@depauw.edu

where, β_0 is a scale parameter and β is a $p \times 1$ vector of parameters. The maximum likelihood estimation procedure is used to estimate the unknown parameters for the model (Hosmer and Lameshow, 2000). Since the logistic model is nonlinear in parameters, an iterative procedure such as Newton-Raphson method is applied (McCullagh and Nelder, 1989). Givens and Hoeting (2005) showed that if $l'''(\beta)$ is continuous and β^* is a simple root of $l'(\beta)$, then there exists a neighborhood of β^* for which Newton-Raphson method converges to β^* when started from any $\beta^{(t)}, t = 1, 2, ...$ in that neighborhood, where $l(\cdot)$ is the log-likelihood of the function. Consistency of the maximum likelihood estimators for logistic regression model was previously studied by different authors, for example, Gourieroux and Monfort (1981), Amemiya (1985). All their work was based upon the fact that the probability of the existence of the estimator $\hat{\beta}$ approaches 1 as n tends to infinity and also assumed that the number of explanatory variable p is compelled to remain constant while sample size n increases. Beer (2001) showed that p is a variable but dependent on n and examined what relationship between p and n is necessary in order not to destroy the consistency of the estimator $\hat{\beta}$. However, this article focuses on a different approach to investigate the consistency of the maximum likelihood estimator $\hat{\beta}$ for the logistic regression model. More precisely, we are going to show that $\hat{\beta}$ converges under certain hypothesis to the real value β^0 if the number of observations (y_i, x_i), where $x_i = (x_{i1,}, \ldots, x_{ip})$, $i = 1, 2, \ldots, n$ tend to infinity. To show this, we follow the procedure described by Lehman and Casella (1998) in which consistency of the maximum likelihood estimators hold if certain regularity conditions are satisfied. It needs to be pointed out that none of the authors (Gourieroux and Monfort, 1981; Amemiya, 1985 and Beer, 2001) verified their work via the Monte Carlo simulation study. Gourieroux and Monfort (1981) noted, "it should be stressed that all these asymptotic results give little indication on the properties of the estimators in finite sample, and it would be interesting to clarify this point by means of Monte Carlo studies." In this article, we provide an extensive standard Monte Carlo simulation study in showing the consistency of the maximum likelihood estimators for the logistic regression model.

This paper is structured as follows. In Section 2, we provide consistency of the maximum likelihood estimators as described by Lehmann and Casella (1998). In section 3, we verified all the conditions needed for consistency noticed in Section

2 for the logistic regression model. A simulation study is presented in Section 4 to demonstrate the performance of consistency for the maximum likelihood estimators. Finally, a concluding remark is given in Section 5.

2. Consistency of the Maximum Likelihood Estimator (MLE)

Lehmann and Casella (1998) provided the following results of the consistency of MLE under some regularity conditions. These are:

(A0) The distributions P_θ of the observations are distinct (otherwise, θ cannot be estimated consistently).

(A1) The distributions P_θ have common support.

(A2) The random variables are $X_i = (X_{i1}, \ldots, X_{ip})$, $i = 1, \ldots, n$ where the X_i's are independent and identically distributed (iid) with probability density $f(x_i \mid \theta)$ with respect to probability measure μ.

(A3) There exists an open subset ω of Ω containing the true parameter point θ^0 such that for almost all x, the density $f(x \mid \theta)$ admits all third derivatives

$$\frac{\partial^3}{\partial\theta_j \partial\theta_k \partial\theta_l} f(x \mid \theta) \text{ for all } \theta \in \omega.$$

(A4) The first and second derivatives of log f satisfy the equations

$$E_\theta\left[\frac{\partial}{\partial\theta_j} \log f(X \mid \theta)\right] = 0 \text{ for } j = 1, \ldots, p, \text{ and}$$

$$I_{jk} = E_\theta\left[\frac{\partial}{\partial\theta_j} \log f(X \mid \theta) \cdot \frac{\partial}{\partial\theta_k} \log f(X \mid \theta)\right] = E_\theta\left[-\frac{\partial^2}{\partial\theta_j \partial\theta_k} \log f(X \mid \theta)\right]$$

(A5) Since the $p \times p$ matrix $I(\theta)$ is a covariance matrix, it is positive semidefinite. We will assume that $I_{jk}(\theta)$ are finite and that the matrix $I(\theta) = ((I_{jk}))$, $j, k = 1, 2, \ldots, p$ is positive definite for all θ in ω, and the p Statistics

$$\frac{\partial}{\partial\theta_1}\log f(x|\theta),\ldots,\frac{\partial}{\partial\theta_p}\log f(x|\theta)$$

are affinely independent with probability 1.

(A6) Finally, we will assume that there exists function M_{jkl} such that

$$\left|\frac{\partial^3}{\partial\theta_j\partial\theta_k\partial\theta_l}\log f(x|\theta)\right| \le M_{jkl}(x) \text{ for all } \theta\in\omega$$

where $m_{jkl} = E_{\theta^0}[M_{jkl}(X)] < \infty$ for all $j,k,l.$

Theorem 1*:* Let $X_1,\ldots,X_n$ be iid each with a density $f(x|\theta)$ (with respect to μ) which satisfies (A0)-(A6) above. Then, with probability tending to 1 as $n\to\infty$, there exist solutions $\hat{\theta}_n = \hat{\theta}_n(X_1,\ldots,X_n)$ of the likelihood equations

$$\frac{\partial}{\partial\theta_j}\left[f(x_1|\theta)\ldots f(x_n|\theta)\right]=0,\ j=1,\ldots,p,$$

or, equivalently,

$$\frac{\partial}{\partial\theta_j}\left[\log L(\theta)\right]=0,\ \ j=1,\ldots,p,$$

such that

(a) $\hat{\theta}_{jn}$ is consistent for estimating θ_j.

(b) is asymptotically normal with mean (vector) zero and covariance matrix $[I(\theta)^{-1}]$, and

(c) $\hat{\theta}_{jn}$ is asymptotically efficient in the sense that

$$\sqrt{n}(\hat{\theta}_{jn}-\theta_j)\xrightarrow{L} N\left\{0,\left[I(\theta)\right]_{jj}^{-1}\right\}.$$

3. Consistency of the MLE in Logistic Model

We verify all the regularity conditions under the logistic regression model discussed in section 3 and then we apply Theorem 1 to show the consistency of MLE for the logistic regression model.

Assumption (A0): Let $\theta_1 = (\beta_0^{(1)}, \beta_1^{(1)}, ..., \beta_p^{(1)})$ and $\theta_2 = (\beta_0^{(2)}, \beta_1^{(2)}, ..., \beta_p^{(2)})$. We define the models as

$$P_{\theta_1}(Y=1 \mid X=x) = \frac{e^{\beta_0^{(1)}x_0 + \beta_1^{(1)}x_1 + \cdots + \beta_p^{(1)}x_p}}{1+e^{\beta_0^{(1)}x_0 + \beta_1^{(1)}x_1 + \cdots + \beta_p^{(1)}x_p}} = \frac{e^{\beta^{(1)T}x}}{1+e^{\beta^{(1)T}x}}$$

$$P_{\theta_2}(Y=1 \mid X=x) = \frac{e^{\beta_0^{(2)}x_0 + \beta_1^{(2)}x_1 + \cdots + \beta_p^{(2)}x_p}}{1+e^{\beta_0^{(2)}x_0 + \beta_1^{(2)}x_1 + \cdots + \beta_p^{(2)}x_p}} = \frac{e^{\beta^{(2)T}x}}{1+e^{\beta^{(2)T}x}}$$

where $\beta^{(1)^T} = (\beta_0^{(1)}, \beta_1^{(1)}, ..., \beta_p^{(1)})$, $\beta^{(2)^T} = (\beta_0^{(2)}, \beta_1^{(2)}, ..., \beta_p^{(2)})$ and $x^T = (x_0, x_1, ..., x_p)$

If $\beta^{(1)} = \beta^{(2)}$, then the above two equations are the same. On the contrary, we are going to show that if the equations are equal, then $\beta^{(1)} = \beta^{(2)}$.

We have, $\dfrac{e^{\beta^{(1)T}x}}{1+e^{\beta^{(1)T}x}} = \dfrac{e^{\beta^{(2)T}x}}{1+e^{\beta^{(2)T}x}}$

This implies, $(\beta_0^{(1)} - \beta_0^{(2)})x_0 + (\beta_1^{(1)} - \beta_1^{(2)})x_1 +, ..., +(\beta_p^{(1)} - \beta_p^{(2)})x_p = 0$

That is, $a_0x_0 + a_1x_1 +, ..., + a_px_p = 0$, where $a_i = \beta_i^{(1)} - \beta_i^{(2)}, \quad i = 0,1,..., p$

Since x_i 's are independent, so $a_0 = a_1 = ... = a_p = 0$, this implies that, $\beta^{(1)} = \beta^{(2)}$.

This indicates that the distributions are unique, therefore, if $\theta_1 \neq \theta_2$, then the distributions P_θ of the observations are distinct.

Assumption (A1): The variables in the model are $x_1, x_2, ..., x_p$, let $x = (x_1, x_2, ..., x_p)$ where $x \in \Re^p$ and the parameter β takes values $-\infty < \beta_j < \infty$, *j*=1,2,…,*p*. This is true for each model stated in the assumption (A0). Therefore, the distributions P_θ have common support.

Assumption (A2): In the logistic model, we consider the observations of the form $x_i = (x_{i1}, \ldots, x_{ip})$, i =1,…,*n* where the x_i are iid with probability density $P(Y=1 \mid x)$.

Assumption (A3): When *Y*=1, we define $f(x \mid \beta) = \dfrac{e^{\beta_0 + \beta^T x_i}}{1+e^{\beta_0 + \beta^T x_i}}$ have the likelihood for the logistic model is proportional to

$$L = \prod_{i=1}^{n} \frac{e^{\beta_0 + \beta^T x_i}}{1 + e^{\beta_0 + \beta^T x_i}}$$

Taking *log* on both sides and we get

$$\log L = \sum_{i=1}^{n} \left[\beta_0 + \beta^T x_i - \log(1 + e^{\beta_0 + \beta^T x_i}) \right]$$

Now, taking derivative with respect to β_j, we have

$$\frac{\delta \log L}{\delta \beta_j} = \sum_{i=1}^{n} \left[x_{ij} - \frac{x_{ij}\, e^{\beta_0 + \beta^T x_i}}{1 + e^{\beta_0 + \beta^T x_i}} \right] = \sum_{i=1}^{n} \left[\frac{x_{ij}}{1 + e^{\beta_0 + \beta^T x_i}} \right]$$

Now the derivative comes to the form $\frac{x_{ij}}{1 + e^{\beta_0 + \beta^T x_i}}$. If we take the derivative of *k*th order, the derivative comes to the form $\frac{x_{ij}}{(1 + e^{\beta_0 + \beta^T x_i})^k}$, which can be proved by the mathematical induction. Therefore, not only does the derivative of $f(x \mid \beta)$ third order exist, but the derivatives of all orders exist.

Assumption (A4): The condition (A4) is proved, in general, for the density $f(x \mid \beta)$ under the condition that the differentiation under the integral sign is allowed. The only thing we need to check for the logistic model is that whether it permits the differentiation under the integral sign. To show that part we consider the following theorem, which is available in standard real analysis or probability books (see, Durrett , 2005). This theorem allows us to perform the differentiation under the integral sign.

Theorem 2. Suppose we are given the following:

- An open interval $I \subset \Re$.
- A measurable subset $X \subset \Re$.
- A function $H : I \times X \to \Re$
- A function $g : X \to [0, \infty]$

Assume the following:

- $\left| \frac{\partial H}{\partial t}(t, x) \right| \leq g(x)$ for every $t \in I$ and $x \in X$.
- g is integrable.

- $t \to H(t,x)$ is a differentiable function of $t \in I$ for every $x \in X$.
- $x \to H(t,x)$ is an integrable function of $x \in X$ for every $t \in I$.

Then the following hold:

- $x \to \frac{\partial H}{\partial t}(t,x)$ is an integrable function of $x \in X$ for every $t \in I$.
- $t \to \int_X H(t,x)dx$ is a differentiable function of $t \in I$.
- $\frac{d}{dt}\int_X H(t,x)dx = \int_X \frac{\partial}{\partial t} H(t,x)dx$ for every $t \in I$.

To verify the above assumptions of Theorem 2 for logistic regression model, we consider the following function when y =1.

$$H(\beta,x) = \frac{e^{\beta_0+\beta^T x}}{1+e^{\beta_0+\beta^T x}}$$

$$\frac{\partial}{\partial \beta} H(\beta,x) = \frac{x_i e^{\beta_0+\beta^T x}}{(1+e^{\beta_0+\beta^T x})^2}$$

$$\left| \frac{\partial}{\partial \beta} H(\beta,x) \right| = \left| \frac{x_i e^{\beta_0+\beta^T x}}{(1+e^{\beta_0+\beta^T x})^2} \right| \le |x| \left| \frac{e^{\beta_0+\beta^T x}}{(1+e^{\beta_0+\beta^T x})^2} \right| = g(x) \text{ as } \left| \frac{e^{\beta_0+\beta^T x}}{(1+e^{\beta_0+\beta^T x})^2} \right| < 1$$

Similarly, this can be shown for y =0.

Since $H(\beta,x)$ is a differentiable function of $x \in X$ for every $\beta \in \Re^p$ and integrable for $x \in X$ for every $\beta \in \Re^p$. Thus the results of the Theorem 2 hold.

Assumption (A5): We take the derivative of $\log f(x|\beta)$ with respect to $\beta_1, \beta_2, \cdots, \beta_p$, we have

$$\frac{\partial \log f}{\partial \beta_j} = x_j - \frac{x_j e^{\beta_0+\beta^T x}}{1+e^{\beta_0+\beta^T x}}, \; j = 1,2,\ldots,p$$

Now we write the vectors in the form so that they are linearly dependent in the following way,

$$x_p - \frac{x_p e^{\beta_0+\beta^T x}}{1+e^{\beta_0+\beta^T x}} = \sum_{j=1}^{p-1}\alpha_j x_j - \sum_{j=1}^{p-1}\alpha_j \frac{x_j e^{\beta_0+\beta^T x}}{1+e^{\beta_0+\beta^T x}}$$

That is, $\left(x_p - \sum_{j=1}^{p-1}\alpha_j x_j\right) - \frac{e^{\beta_0+\beta^T x}}{1+e^{\beta_0+\beta^T x}}\left(x_p - \sum_{j=1}^{p-1}\alpha_j x_j\right) = 0$

That is, $\left(x_p - \sum_{j=1}^{p-1}\alpha_j x_j\right)\left(1 - \frac{e^{\beta_0+\beta^T x}}{1+e^{\beta_0+\beta^T x}}\right) = 0 \quad \forall\ \beta$

Since $\left(1 - \frac{e^{\beta_0+\beta^T x}}{1+e^{\beta_0+\beta^T x}}\right) \neq 0$, So, $\left(x_p - \sum_{j=1}^{p-1}\alpha_j x_j\right) = 0$

Thus, $x_p = \sum_{j=1}^{p-1}\alpha_j x_j$

We have, $P_\beta\left[x_p = \sum_{j=1}^{p-1}\alpha_j x_j\right] = 0, \quad \forall\ \beta$ because the joint distribution of $x_1, x_2, ..., x_n$ is continuous on $\Re^p$.

Thus, $P_\beta\left[x_p \neq \sum_{j=1}^{p-1}\alpha_j x_j\right] = 1, \quad \forall\ \beta$

This implies that the statistics are affinely independent.

Assumption (A6): We have, $\frac{\delta \log L}{\delta \beta_j} = \sum_{i=1}^{n}\left[x_{ij} - \frac{x_{ij}\ e^{\beta_0+\beta^T x}}{1+e^{\beta_0+\beta^T x}}\right] = \sum_{i=1}^{n}\left[\frac{x_{ij}}{1+e^{\beta_0+\beta^T x}}\right]$

and $\frac{\delta^2 \log L}{\delta \beta_j \delta \beta_k} = -\sum_{i=1}^{n} x_{ij} x_{ik}\left[\frac{e^{\beta_0+\beta^T x}}{\left(1+e^{\beta_0+\beta^T x}\right)^2}\right]$

So, $\frac{\delta^3 \log L}{\delta\beta_j\ \delta\beta_k\ \delta\beta_l} = -\sum_{i=1}^{n} x_{ij} x_{ik} x_{il}\left(\frac{e^{\beta_0+\beta^T x}(1-e^{\beta_0+\beta^T x})}{(1+e^{\beta_0+\beta^T x})^3}\right)$

$$\therefore \quad \left|\frac{\delta^3 \log L}{\delta\beta_j\ \delta\beta_k\ \delta\beta_l}\right| = \left|-\sum_{i=1}^{n} x_{ij} x_{ik} x_{il}\left(\frac{e^{\beta_0+\beta^T x}(1-e^{\beta_0+\beta^T x})}{(1+e^{\beta_0+\beta^T x})^3}\right)\right|$$

$$\leq \sum_{i=1}^{n} |x_{ij}x_{ik}x_{il}| \left| \frac{e^{\beta_0+\beta^T x}(1-e^{\beta_0+\beta^T x})}{(1+e^{\beta_0+\beta^T x})^3} \right|$$

$$\leq \sum_{i=1}^{n} |x_{ij}x_{ik}x_{il}| \quad \text{since} \left| \frac{e^{\beta_0+\beta^T x}(1-e^{\beta_0+\beta^T x})}{(1+e^{\beta_0+\beta^T x})^3} \right| < 1$$

$$\leq \sum_{i=1}^{n} |x_{ij}|\,|x_{ik}|\,|x_{il}|$$

$$\leq M_{jkl}(x)$$

where $m_{jkl} = E[M_{jkl}(X)] = E\left[\sum_{i=1}^{n}|X_{ij}|\,|X_{ik}|\,|X_{il}|\right] = \sum_{i=1}^{n} E|X_{ij}X_{ik}X_{il}|$, which is finite. Since the logistic model satisfies all the regularity conditions (A0)-(A6), therefore, $\hat{\beta}$ converges to the real value β_0 by Theorem 1.

4. A Simulation Study

We now assess, via standard Monte Carlo simulation, the finite sample performance of consistency of the maximum likelihood estimators. In the simulation study, we consider four explanatory variables x_1, x_2, x_3, and x_4 which are fixed and the binary response variable *y*, which is treated as a random variable in the logistic regression model. For the fixed values of the intercept parameter β_0 and four other parameters β_1, β_2, β_3, and β_4, our aim is to compare the performance of the values of parameters and their standard errors when sample size increases. For fixed values of β_0 = 0.7, β_1=1.0, β_2=1.3, β_3 =0.25, and β_4=0.05, the logistic regression model becomes:

$$\pi(x) = \frac{e^{0.7+1.0\,x_i+1.3\,x_2+0.25\,x_3+0.05x_4}}{1+e^{0.7+1.0\,x_i+1.3\,x_2+0.25\,x_3+0.05x_4}}$$

In the simulation, we consider sample sizes of $n=50$, 100, 150, and 200 and generate 1,000 independent sets of random samples for each different sample size. For each set of random sample with particular sample size, we estimate β_0, β_1, β_2, β_3 and β_4 and their standard errors based on the logistic regression model. The final estimates and standard errors of β_0, β_1, β_2, β_3, and β_4 are the average of 1,000 estimates of β_0, β_1, β_2, β_3, and β_4 for that particular sample size. The following Table gives the results of simulation study for different sample sizes.

Table 1: Estimated parameter values and their standard errors using the logistic regression model for different sample sizes of 50, 100, 150, and 200

Parameters	$n=50$		$n=100$		$n=150$		$n=200$	
	Estimate	SE	Estimat	SE	Estimate	SE	Estimate	SE
β_0	1.23556	0.131	0.86350	0.0426	0.73616	0.0165	0.74692	0.0147
β_1	2.64360	0.184	1.26267	0.0575	1.08362	0.0258	1.08228	0.0249
β_2	4.14250	0.225	1.75881	0.0808	1.46068	0.0414	1.38185	0.0248
β_3	1.02962	0.158	0.31986	0.0413	0.25220	0.0166	0.26272	0.0147
β_4	0.37990	0.147	0.01565	0.0442	0.06005	0.0173	0.04498	0.0148

SE=Standard Error

As seen in the above Table, for sample size $n=50$, the estimated values of parameters are different from the true values (β_0 = 0.7, β_1=1.0, β_2=1.3, β_3=0.25, and β_4=0.05), and also the standard errors become large. However, when the sample size increases from $n=50$ to $n=200$, the estimated values of the parameters β_0, β_1, β_2, β_3, and β_4 are close to the true values, and standard errors of the estimates are noticeably smaller. This indicates that simulation study performs well in showing the consistency of maximum likelihood estimators for parameters of the logistic regression model.

5. Conclusion

This paper investigates a different approach to show the consistency of maximum likelihood estimators in the logistic regression model. In that approach, we verify all the regularity conditions for consistency mentioned in section 2 for the logistic regression model and conclude that the parameters of the logistic regression model converge to its true values when sample size increases. This paper also concentrates on Monte Carlo simulation study for showing the consistency of the maximum likelihood estimators for the logistic regression model. Results indicate that the simulation study performs very well and the simulation standard errors of the parameters get smaller as the sample size increases.

References

1. Amemiya, T. (1985). *Advanced Econometrics*. Harvard University Press, Cambridge.
2. Beer, M. (2001). Asymptotic properties of the maximum likelihood estimator in dichotomous logistic regeression models. *Diploma Thesis in Mathematics*. University of Fribourg Switzerland.
3. Durrett, R. (2005). *Probability: Theory and Exa*mples. Third Ed., Thomson, Brooks/Cole.
4. Givens, G. H. and Hoeting, A. J. (2005). *Computational Statistics*. Wiley Interscience.
5. Gourieroux, C. and Monfort, A. (1981). Asymptotic properties of the maximum likelihood estmator in dichotomous logit models. *Journal of Econometrics*, **17(1)**, 83-97.
6. Hosmer, D. W. and Lameshow, S. (2000). *Applied Logistic Regression*. John Wiley and Sons, New York.
7. Lehmann, E. L. and Casella,G. (1998). *Theory of Point Estimation*. Second, Ed. Springer, New York.
8. McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. London Chapman and Hall.
9. Mehta, C. R., Patel, N. R. and Senchaudhuri, P. (2000). Efficient Monte Carlo methods for conditional logistic regression. *Journal of the American Statistical Association*, **95**, 99-108.