Journal of Mathematics and Statistics 4 (2): 108-111, 2008
ISSN 1549-3644

The Estimation of Logistic Probability Model of Binomial Data: Simulation Comparison

Zaki. J. Al- Sarraf, Faris M. Al- Athari and Hilmi F. Kittani
Department of Mathematics, Hashemite University, P.O. Box 150459,
Zarqa 13115, Jordan

Abstract: Properties of various types of estimators of the regression coefficients in linear logistic regression models were considered. The estimators include those based on maximum likelihood, minimum chi-square and weighted least squares. The results of a large scale simulation investigation evaluating the moment properties of the estimators are presented for the case of logistic model with a single explanatory variable.

Keywords: Maximum likelihood, newton- raphson approach, weighted least squares

INTRODUCTION

Assume that y_1, y_2,...,y_g represent g independent binomial random variables. For ith group, i = 1, 2,…, g, let x_{i1},x_{i2},...,x_{ik} denote the values on k explanatory variable which are thought to influence the individual trial probability of success, denoted by P_i. Then the linear logistic regression model is:

$$\ln(P_i / Q_i) = \mathbf{x}_i' \beta \quad i = 1, 2, \ldots, g \tag{1}$$

where $Q_i = 1 - P_i$,

$$x_i' = (1, x_{i1}, \ldots, x_{ik}), \text{and } \beta' = (\beta_0, \beta_1, \ldots, \beta_k) \tag{2}$$

The regression coefficients in β are usually unknown and there are a number of well-known methods of estimation of β, such as maximum likelihood, minimum chi- square and weighted least squares.

Maximum likelihood: The maximum likelihood (ML)[1-4], is the most common method of estimation, since these estimates can now be routinely obtained by using many statistical packages such as GLIM, SHAZAM, MLOGIT, QUAIL, SAS PROC GENMOD and S-PLUS. Refer to[5-7] for more details.

The kernel of the log-likelihood may be written a:

$$L(\beta) = \sum_{i=1}^{g} n_i (p_i x_i' \beta' - \ln(1 + e^{x_i' \beta})) \tag{3}$$

where $p_i = y_i/n_i$ denotes the observed proportion of successes in the i^{th} group. In a matrix form the first and second order derivatives of the log- likelihood are given by:

$$\frac{\partial L(\beta)}{\partial(\beta)} = \begin{pmatrix} \sum_{i=1}^{g} n_i x_{i1} (p_i - P_i) \\ \sum_{i=1}^{g} n_i x_{i2} (p_i - P_i) \\ \vdots \\ \sum_{i=1}^{g} n_i x_{ik} (p_i - P_i) \end{pmatrix} \tag{4}$$

and

$$\frac{\partial^2 L(\beta)}{\partial(\beta)\partial(\beta')} = -X' V_1 X \tag{5}$$

where v1 = diag $((n_i P_i Q_i))$ is a (g x g) diagonal matrix and X is the (g x (k+1) design matrix[2] . If we let $\hat{\beta}_1$ denote the ML estimate of β and put $D_1 = (\frac{\partial L(\beta)}{\partial \beta})_{\beta = \hat{\beta}_1}$, $\hat{V}_1 = (V_1)_{\beta = \hat{\beta}_1}$, then $\hat{\beta}_1$ is given by the solution of the (k+1) equations given by $D_1 = 0$.

Applying a standard Newton-Raphson approach[8] to solve these equations, if $\hat{\beta}_1^{(\ell)}$ denotes the approximation to $\hat{\beta}_1$ at the lth stage of iteration we have:

$$\hat{\beta}_1^{(\ell+1)} = \hat{\beta}_1^{(\ell)} + (X' \hat{V}_1^{(\ell)} X)^{-1} D_1^{(\ell)} \tag{6}$$

Corresponding Author: Hilmi F. Kittani, Department of Mathematics, Hashemite University, PO Box 150459, Zarqa13115 Zarqa, Jordan

where $\hat{V}_1^{(\ell)}$ and $D_1^{(\ell)}$ denotes $\hat{V}_1$ and D_1 evaluated at $\hat{\beta}_1^{(\ell)}$. The procedure is to calculate the estimates from a linear probability model and to use these as an initial values which begins finding a solution. When the difference between $\hat{\beta}_1^{(\ell+1)}$ and $\hat{\beta}_1^{(\ell)}$ is close enough to zero, the process stops. This iteration procedure may be viewed as a method of reweighted least squares.

Minimum chi- squares: The minimum chi- square (MCS) estimator[2,4] which we denote by $\hat{\beta}_2$ is the value of β that minimizes:

$$R(\beta) = \sum_{i=1}^{g} \frac{n_i (p_i - P_i)^2}{P_i Q_i} \tag{7}$$

Writing

$$R(\beta) = \sum_{i=1}^{g} n_i (p_i^2 \frac{Q_i}{P_i} + q_i^2 \frac{P_i}{Q_i} - 2p_i q_{i)} \tag{8}$$

$$\sum_{i=1}^{g} n_i (p_i^2 e^{-x_i' \beta} + q_i^2 e^{x_i' \beta} - 2p_i q_i) \tag{9}$$

The first and the second order derivatives of R (β) are:

$$\frac{\partial R(\beta)}{\partial \beta} = \begin{pmatrix} \sum n_i x_{i0} (q_i^2 \frac{P_i}{Q_i} - p_i^2 \frac{Q_i}{P_i}) \\ \sum n_i x_{i1} (q_i^2 \frac{P_i}{Q_i} - p_i^2 \frac{Q_i}{P_i}) \\ \vdots \\ \sum n_i x_{ik} (q_i^2 \frac{P_i}{Q_i} - p_i^2 \frac{Q_i}{P_i}) \end{pmatrix} \tag{10}$$

$$\frac{\partial^2 R(\beta)}{\partial \beta \partial \beta'} = X' V_2 X \tag{11}$$

Where;

$$V_2 = \text{diag}((n_i (p_i^2 \frac{Q_i}{P_i} - q_i^2 \frac{P_i}{Q_i})) \tag{12}$$

Then $\hat{\beta}_2$ is given by the solution of the (k+1) equations given by:

$$D_2 = 0 \tag{13}$$

An iterative solution can again be found by using Newton- Raphson approach similar to that outlined for the maximum likelihood estimation procedure.

Weighted least squares: The ML and MCS estimation methods both require an iteration method of solution, although as we have noted the estimates can be obtained using many packages (straight forward). A none iterative solution can be found by using weighted least squares (WLS), which is sometimes referred to as minimum logit chi- square estimation[9]. Defining the group empirical logits by:

$$z_i = \ln(\frac{p_i}{q_i}),\ i = 1,2,...,g \tag{14}$$

we have,

$$E(z_i) = x_i' \beta + o(n_i^{-1}) \tag{15}$$

and

$$v(z_i) = (n_i P_i Q_i)^{-1} + o(n_i^{-2}) \tag{16}$$

The WLS estimate is the value of β which minimizes:

$$S(\beta) = \sum_{i=1}^{g} n_i p_i q_i (\ln \frac{p_i}{q_i} - \ln \frac{P_i}{Q_i})^2 \tag{17}$$

$$\sum_{i=1}^{g} w_i (z_i - x_i' \beta)^2 \tag{18}$$

The weights $w_i = n_i p_i q_i$ are the reciprocal of the asymptotic variances of the z_i's based on the sample proportions of successes. Denoting the WLS estimate by $\hat{\beta}_3$, the explicit solution is:

$$\hat{\beta}_3 = (X'WX)^{-1} X'Wz \tag{19}$$

where $\mathbf{z}' = (z_1, z_2, ..., z_g)$ and $W = \text{diag}((w_i))$.

The WLS method can be applied even if p = 0 or 1 but because z_i is itself undefined for these extreme cases therefore samples with $y_i = 0$ or n_i should be exempted when the estimation of $\hat{\beta}_3$ is considered.

We use $\hat{\beta}$ to denote any estimator from the set $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$. It follows that the three estimators all have the same asymptotic properties with $E(\hat{\beta}) = \beta$ and:

$$COV(\hat{\beta}) = (X' V_1 X)^{-1} \tag{20}$$

In the results of a fairly large scale simulation investigation that compare the moment properties of the estimators for a number of sample sizes and parameter configurations when we consider a single explanatory variable. These results considerably extend the findings made by[10] who considered the particular case g = 3, n_i = 1,2,3 and showed that the simple WLS method was more efficient than the ML and MCS methods of estimation under a number of success probability configurations.

Numerical results with discussions: In order to investigate the properties of the ML, MCS an WLS estimators, a large scale simulation was made for the case of a single explanatory variable with equally spaced values. Without loss of generality, the linear logistic regression model was taken as:

$$\mathrm{Ln}(P_i / Q_i) = \beta_0 + \beta_1(i-1),\ i = 1,2,\ldots,g \quad (21)$$

Group numbers g = 5, 10 and sample sizes n = 25, 50, 100 were used, the simulation run size being, 1000 in each case. Three pairs of values (β_0, β_1) were examined for each value of g to give coverage of markedly different configurations for the groups trial probabilities of success, as shown in Table 1.

For the model given by Eq. 21, the elements in the information matrix are:

$$\mathrm{Ln}(P_i / Q_i) = \beta_0 + \beta_1(i-1),\ i = 1,2,\ldots,g \quad (22)$$

where, x_i = i-1The asymptotic variances and covariances for each of the three estimators are given by:

$$\mathrm{var}(\hat{\beta}_0) = I_{22/D},\ \mathrm{var}(\hat{\beta}_1) = I_{11} / D,$$
$$\mathrm{cov}\ (\hat{\beta}_0, \hat{\beta}_1) = -I_{12} / D \quad (23)$$

where $D = I_{11} I_{22} - I_{12}^2$.

In Table 2, the variances of the estimators of β_0 and β_1 as obtained by simulation are given together with the approximate values given by 23. The results show that the ML estimators gave the largest variances for both β_0 and β_1 in all cases. The variance differences for the MCS and WLS estimators were generally extremely small. The results also show that approximation 23 which is applicable to all the estimators gave satisfactory results especially when n = 100.

Table 3 gives the means square errors of the estimators and the efficiencies of the MCS and WLS estimators relative to ML estimator defined as the ratio of the mean square errors. These efficiency results suggest that the ML estimation procedure should not be

Table 1: Parameter values (β_0,β_1) and group trial probabilities of success {P_i} used in the simulation investigation

	β_0	β_1	{P_i}
g = 5			
(i)	-2.0	0.4	0.119 0.168 0.232 0.310 0.401
(ii)	-1.0	0.5	0.269 0.378 0.500 0.623 0.731
(iii)	0.5	0.5	0.623 0.731 0.818 0.881 0.924
g = 10			
(iv)	-2.0	2.0	0.119 0.142 0.168 0.198 0.231 0.269 0.310 0.354 0.401 0.450
(v)	-0.4	0.2	0.401 0.450 0.500 0.550 0.591 0.646 0.690 0.731 0.769 0.802
(vi)	0.5	0.2	0.623 0.668 0.711 0.750 0.785 0.818 0.846 0.870 0.891 0.908

Table 2: Variances of estimators for configurations shown in Table 1

Configuration		ML	MCS	WLS	Approx (23)
a) β_0					
	(i)	0.2118	0.1848	0.1768	0.1889
	(ii)	0.1143	0.1083	0.1070	0.1143
n = 25	(iii)	0.1272	0.1187	0.1190	0.1176
	(iv)	0.1127	0.0998	0.0983	0.1018
	(v)	0.0603	0.0560	0.0552	0.0586
	(vi)	0.0661	0.0617	0.0620	0.0688
	(i)	0.1017	0.0957	0.0940	0.0944
	(ii)	0.0597	0.0581	0.0579	0.0571
n = 50	(iii)	0.0627	0.0607	0.0605	0.0588
	(iv)	0.0538	0.0513	0.0510	0.0509
	(v)	0.0282	0.0273	0.0273	0.0293
	(vi)	0.0383	0.0372	0.0370	0.0344
	(i)	0.0480	0.0464	0.0459	0.0472
	(ii)	0.0301	0.0297	0.0296	0.0286
n = 100	(iii)	0.0319	0.0316	0.0315	0.0294
	(iv)	0.0270	0.0264	0.0264	0.0255
	(v)	0.0141	0.0139	0.0139	0.0146
	(vi)	0.0174	0.0172	0.0172	0.0172
b) β_1 (var iances×10^2)					
	(i)	2.5560	2.2950	2.2240	2.5013
	(ii)	1.8980	1.7930	1.7740	1.9580
n = 25	(iii)	3.7530	3.1790	3.2030	3.1078
	(iv)	0.3037	0.2721	0.2668	0.2903
	(v)	0.2476	0.2278	0.2234	0.2360
	(vi)	0.3098	0.2758	0.2761	0.3386
	(i)	1.3770	1.3120	1.2900	1.2506
	(ii)	1.0670	1.0370	1.0320	0.9790
n = 50	(iii)	1.6980	1.6020	1.5880	1.5539
	(iv)	0.1548	0.1491	0.1486	0.1451
	(v)	0.1197	0.1144	0.1138	0.1180
	(vi)	0.1804	0.1718	0.1700	0.1693
	(i)	0.5923	0.5762	0.5711	0.6253
	(ii)	0.4945	0.4874	0.4863	0.4895
n = 100	(iii)	0.8168	0.7977	0.7942	0.7770
	(iv)	0.0750	0.0740	0.0739	0.0726
	(v)	0.0593	0.0585	0.0583	0.0590
	(vi)	0.0865	0.0854	0.0849	0.0846

used when the groups sample sizes are less than 100. The efficiency differences between the MCS and WLS procedures are consistently small and WLS has the computational advantage of providing a non-iterative fit.

Table 3: Mean square errors and percentage efficiencies of estimators relative to the ML estimators for configurations shown in Table 1

Configuration		Mean square errors		Efficiencies		
a) β_0						
	ML	MCS	WLS	MCS	WLS	
	(i)	0.2205	0.1851	0.1772	119.1	124.4
	(ii)	0.1161	0.1087	0.1072	106.8	108.3
n = 25	(iii)	0.1277	0.1190	0.1197	107.3	106.7
	(iv)	0.1135	0.1043	0.1079	108.8	105.2
	(v)	0.0604	0.0560	0.0552	107.9	109.5
	(vi)	0.0661	0.0618	0.0620	107.0	106.6
	(i)	0.1027	0.0957	0.0942	107.3	109.0
	(ii)	0.0597	0.0582	0.0580	102.6	102.9
n = 50	(iii)	0.0627	0.0607	0.0605	103.3	103.6
	(iv)	0.0538	0.0531	0.0536	101.3	100.4
	(v)	0.0284	0.0273	0.0273	104.0	104.0
	(vi)	0.0383	0.0372	0.0371	103.0	103.2
	(i)	0.0482	0.0464	0.0460	103.9	104.8
	(ii)	0.0302	0.0297	0.0296	101.7	102.0
n = 100	(iii)	0.0319	0.0316	0.0315	101.0	101.3
	(iv)	0.0270	0.0269	0.0270	100.4	100.0
	(v)	0.0141	0.0139	0.0137	101.4	101.4
	(vi)	0.0174	0.0172	0.0172	101.2	101.2
b) β_1 (mean square errors$\times 10^2$)						
	(i)	2.6311	2.3040	2.2240	114.2	118.3
	(ii)	1.9311	1.7970	1.7752	107.5	108.8
n = 25	(iii)	3.7684	3.2160	3.3745	117.2	111.7
	(iv)	0.3051	0.2767	0.2783	110.3	109.6
	(v)	0.2491	0.2288	0.2255	108.9	110.5
	(vi)	0.3099	0.2897	0.3099	107.0	100.0
	(i)	1.3879	1.3120	1.2900	105.8	107.6
	(ii)	1.0689	1.0370	1.0323	103.1	103.6
n = 50	(iii)	1.7040	1.6060	1.6028	106.1	106.3
	(iv)	0.1548	0.1512	0.1518	102.4	102.0
	(v)	0.1205	0.1144	0.1140	105.3	105.7
	(vi)	0.1805	0.1744	0.1749	103.5	103.2
	(i)	0.5936	0.5762	0.5713	103.0	103.9
	(ii)	0.4951	0.4875	0.4863	101.6	101.8
n = 100	(iii)	0.8265	0.7986	0.7944	103.5	104.0
	(iv)	0.0750	0.0746	0.0746	100.5	100.5
	(v)	0.0593	0.0588	0.0587	100.9	101.0
	(vi)	0.0867	0.0856	0.0852	101.3	101.8

REFERENCES

1. Montgomery D.C., Peck, E.A. and Vining, G. G. (2001). Introduction to Linear Regression Analysis, 3rd ed., Wiley, New York, , pp.193-196. ISBN 0-471-31565-6
2. Berkson, J. (1995). Maximum likelihood and minimum Chi-square estimates of the logistic function, J. Amer. Statist. Assoc., Vol. 50, pp. 130-162.links.jstor.org/sici?sici=0006-341X(198509)41%3A3%3C777%3AMLVMC%3E2.0.CO%3B2-W
3. Verbeek, M. (2000). A guide to modern Econometrics , 3rd ed. Wiley, New York, pp.152-174. ISBN: 978-0-470-51769-7
4. Hogg, R.V. and Craig, A. T. (1995). Introduction To Mathematical Statistics, 5th ed., Prentice Hall, New Jersey, pp.261-266 & 298-299. ISBN-10: 0023557222
5. Baker, R.J. and Nelder, J.A. (1978). Generalized Linear Interactive Modeling (GLIM), Numerical Algorithms Group, Journal of the American Statistical Association, Vol. 74, No. 368 (Dec., 1979), pp. 934-935. http://www.jstor.org/action/showPublisher?publisherCode=astata
6. Amemiya, T. Qualitative Response Model (1981): A Survey, J. of Economic Literature. Vol.19, pp. 1488-1536.http://www.jstor.org/fcgi-bin/jstor/listjournal.fcg/00220515/.11-.20
7. SAS Institute (1994): Principles of Regression Analysis Course Notes, SAS Institute Publishing company, Cary, N.C. ISBN-10: 1555446043
8. Burden, R L. and Faires, J. D. (2001). Numerical Analysis, 7th ed., Wadsworth Group, pp. 66-72. ISBN-10: 0534382169
9. Johnson, R. A. and Wichern, D., W. (1998). Applied multivariate statistical analysis. 4th ed., Prentice Hall, New Jersey, (1998), pp.187-183, 448. ISBN 0-13-834194-x
10. Johnston, J. and Dinardo, J. (1997). Econometric Methods, 4th ed., McGraw-Hill, New York, pp.342-343,433-434. ISBN: 0070327203