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 \( y_1, y_2,\ldots,y_g \) represent \( g \) independent binomial random variables. For \( i \)th group, \( i = 1, 2,\ldots, g \), let \( x_{i1},x_{i2},\ldots,x_{ik} \) 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:

\[
\ln \left( \frac{P_i}{1-P_i} \right) = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}
\]

where \( P_i = \frac{y_i}{n_i} \) 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:

\[
\frac{\partial L(\beta)}{\partial \beta} = \sum_{i=1}^{g} n_i x_i (p_i - P_i) \quad \frac{\partial^2 L(\beta)}{\partial \beta^2} = -X^T V X
\]

and

\[
\beta = (V X)^{-1} X^T \ln (y_i/n_i)
\]

where \( V = \text{diag}(n_i P_i (1-P_i)) \) is a \((g \times g)\) diagonal matrix and \( X \) is the \((g \times (k+1))\) design matrix. If we let \( \beta \) denote the ML estimate of \( \beta \) and put \( D_i = \frac{\partial^2 L(\beta)}{\partial \beta^2} \) and \( V_i = (V_i)_{ij} \), then \( \beta \) is given by the solution of the \((k+1)\) equations given by \( D_i = 0 \).

Applying a standard Newton-Raphson approach to solve these equations, if \( \beta^{(l)} \) denotes the approximation to \( \beta \) at the \( l \)th stage of iteration we have:

\[
\beta^{(l+1)} = \beta^{(l)} + (X^T V_i X)^{-1} D_i
\]
where $\tilde{V}_l^{(r)}$ and $D_l^{(r)}$ denotes $V_l$ and $D_l$, evaluated at $\hat{\beta}_l^{(r)}$. 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 $\beta_l^{(r)}+\beta_l^{(r)}$ and $\beta_l^{(r)}$ 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 which we denote by $\beta_2$ is the value of $\beta$ that minimizes:

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

Writing

$$R(\beta) = \sum_{i=1}^{g} n_i(p_i^2 + q_i - 2p_iq_i)$$

The first and the second order derivatives of $R(\beta)$ are:

$$\frac{\partial R(\beta)}{\partial \beta} = \sum_{i=1}^{g} n_i(x_i^2 - x_i)$$

$$\frac{\partial^2 R(\beta)}{\partial \beta \partial \beta'} = X'V_x X$$

Where;

$$V_x = \text{diag}(n_i(p_i^2 - q_i)^2)$$

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

$$D_2 = 0$$

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. Defining the group empirical logits by:

$$z_i = \ln\left(\frac{p_i}{q_i}\right), i = 1, 2, \ldots, g$$

we have,

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

and

$$v(z_i) = (n_ip_iq_i)^{-1} + o(n_i^{-2})$$

The WLS estimate is the value of $\beta$ which minimizes:

$$S(\beta) = \sum_{i=1}^{g} n_i(p_i - \ln \left(\frac{p_i}{q_i}\right))^2$$

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

The weights $w_i = n_ip_iq_i$ are the reciprocal of the asymptotic variances of the $z_i$'s based on the sample proportions of successes. Denoting the WLS estimate by $\beta_3$, the explicit solution is:

$$\hat{\beta}_3 = (X'W_X)^{-1}X'W_z$$

where $z' = (z_1, z_2, \ldots, z_g)$ 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 $\beta$ 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 $E(\beta) = \beta$ and:

$$\text{COV}((\beta)) = (X'V_X)^{-1}X'V_X$$
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 and 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:

\[
\ln\left(\frac{P_i}{Q_i}\right) = \beta_0 + \beta_1 (i-1), \quad i = 1,2,\ldots,g
\]  

(21)

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

\[
\text{var}(\hat{\beta}_0) = I_{22}/D, \quad \text{var}(\hat{\beta}_1) = I_{11}/D, \\
\text{cov}(\hat{\beta}_0, \hat{\beta}_1) = -I_{12}/D
\]  

(23)

where \( D = I_{11}I_{22} - I_{12}^2 \).

In Table 2, the variances of the estimators of \( \beta_0 \) and \( \beta_1 \) 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 \( \beta_0 \) and \( \beta_1 \) 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 (\( \beta_0, \beta_1 \)) and group trial probabilities of success \( \{P_i\} \) used in the simulation investigation

| \( g \) | \( \beta_0 \) | \( \beta_1 \) | \( \{P_i\} \) |
|---|---|---|---|
| 5 | (i) -2.0 | 0.4 | \{0.119, 0.168, 0.232, 0.310\} |
| | (ii) -1.0 | 0.5 | \{0.269, 0.378, 0.500, 0.623\} |
| | (iii) 0.5 | 0.5 | \{0.623, 0.731\} |
| 10 | (iv) -2.0 | 2.0 | \{0.119, 0.142, 0.198\} |
| | (v) -0.4 | 0.2 | \{0.401, 0.500\} |
| | (vi) 0.5 | 0.2 | \{0.623, 0.668\} |

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

| Configuration | ML | MCS | WLS | Approx (23) |
|---|---|---|---|---|
| | \( \beta_0 \) | \( \beta_1 \) | \( \{P_i\} \) |
| \( g = 5 \) | | | | |
| n = 25 | (i) 0.2118 | 0.1017 | 0.0597 | 0.1889 |
| | (ii) 0.1143 | 0.0597 | 0.0597 | 0.1143 |
| | (iii) 0.1272 | 0.1017 | 0.0597 | 0.1176 |
| | (iv) 0.1127 | 0.1017 | 0.0597 | 0.1108 |
| | (v) 0.1103 | 0.1017 | 0.0597 | 0.1081 |
| | (vi) 0.1103 | 0.1017 | 0.0597 | 0.1081 |
| \( g = 10 \) | | | | |
| n = 25 | (i) 0.1143 | 0.0597 | 0.0597 | 0.1143 |
| | (ii) 0.1272 | 0.1017 | 0.0597 | 0.1176 |
| | (iii) 0.1127 | 0.1017 | 0.0597 | 0.1108 |
| | (iv) 0.1103 | 0.1017 | 0.0597 | 0.1081 |
| | (v) 0.1103 | 0.1017 | 0.0597 | 0.1081 |
| | (vi) 0.1103 | 0.1017 | 0.0597 | 0.1081 |

**Table 3**: Means square errors of the estimators and efficiencies of the MCS and WLS estimators relative to ML estimator defined as the ratio of the mean square errors

| Configuration | ML | MCS | WLS | Approx (23) |
|---|---|---|---|---|
| | \( \beta_0 \) | \( \beta_1 \) | \( \{P_i\} \) |
| \( g = 5 \) | | | | |
| n = 25 | (i) 2.5560 | 2.2950 | 2.2240 | 2.5013 |
| | (ii) 1.8980 | 1.7930 | 1.7740 | 1.9580 |
| | (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 |
| \( g = 10 \) | | | | |
| n = 25 | (i) 2.5560 | 2.2950 | 2.2240 | 2.5013 |
| | (ii) 1.8980 | 1.7930 | 1.7740 | 1.9580 |
| | (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 |
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 |
|---------------|-------------------|--------------|
|               | ML    | MCS  | WLS  | ML    | MCS  | WLS  |
| a) \( \beta_0 \) |      |      |      |      |      |      |
| n = 25        |      |      |      |      |      |      |
| (i)           | 0.2205 | 0.1851 | 0.1772 | 119.1 | 124.4 |      |
| (ii)          | 0.1161 | 0.1087 | 0.1072 | 106.8 | 108.3 |      |
| (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 |      |
|               | 0.1027 | 0.0957 | 0.0942 | 107.3 | 109.0 |      |
|               | 0.0597 | 0.0582 | 0.0580 | 102.6 | 102.9 |      |
| n = 50        |      |      |      |      |      |      |
| (i)           | 0.0627 | 0.0607 | 0.0605 | 103.3 | 103.6 |      |
| (ii)          | 0.0538 | 0.0531 | 0.0536 | 101.3 | 100.4 |      |
| (iii)         | 0.0284 | 0.0273 | 0.0273 | 104.0 | 104.0 |      |
| (iv)          | 0.0383 | 0.0372 | 0.0371 | 103.0 | 103.2 |      |
| (v)           | 0.0482 | 0.0464 | 0.0460 | 103.9 | 104.8 |      |
| (vi)          | 0.0302 | 0.0297 | 0.0296 | 101.7 | 102.0 |      |
| n = 100       |      |      |      |      |      |      |
| (i)           | 0.0270 | 0.0269 | 0.0270 | 100.4 | 100.0 |      |
| (ii)          | 0.0141 | 0.0139 | 0.0137 | 101.4 | 101.4 |      |
| (iii)         | 0.0174 | 0.0172 | 0.0172 | 101.2 | 101.2 |      |
| b) \( \beta_1 \) (mean square errors×10²) |      |      |      |      |      |      |
| n = 25        |      |      |      |      |      |      |
| (i)           | 2.6311 | 2.3040 | 2.2240 | 114.2 | 118.3 |      |
| (ii)          | 1.9311 | 1.7970 | 1.7752 | 107.5 | 108.8 |      |
| n = 50        |      |      |      |      |      |      |
| (i)           | 3.7684 | 3.2160 | 3.3745 | 117.2 | 111.7 |      |
| (ii)          | 3.051 | 2.767 | 2.783 | 110.3 | 109.6 |      |
| (iii)         | 2.491 | 2.238 | 2.255 | 108.9 | 110.5 |      |
| (iv)          | 0.599 | 0.289 | 0.309 | 107.0 | 100.0 |      |
| (v)           | 1.3879 | 1.3120 | 1.2900 | 105.8 | 107.6 |      |
| (vi)          | 1.0689 | 1.0370 | 1.0323 | 103.1 | 103.6 |      |
| n = 100       |      |      |      |      |      |      |
| (i)           | 1.7040 | 1.6060 | 1.6028 | 106.1 | 106.3 |      |
| (ii)          | 1.548 | 1.512 | 1.518 | 102.4 | 102.0 |      |
| (iii)         | 1.205 | 1.144 | 1.140 | 105.3 | 105.7 |      |
| (iv)          | 1.085 | 1.1744 | 1.035 | 103.5 | 103.2 |      |
| (v)           | 0.5936 | 0.5762 | 0.5713 | 103.0 | 103.9 |      |
| (vi)          | 0.4951 | 0.4875 | 0.4863 | 101.6 | 101.8 |      |
| n = 100       |      |      |      |      |      |      |
| (i)           | 0.8265 | 0.7986 | 0.7944 | 103.5 | 104.0 |      |
| (ii)          | 0.0750 | 0.0746 | 0.0746 | 100.5 | 100.5 |      |
| (iii)         | 0.0593 | 0.0588 | 0.0587 | 100.9 | 101.0 |      |
| (iv)          | 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. link.jstor.org/sici?sici=0006-341X(198509)41%3A3%3C777%3AMLMC%3E2.0.C0%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/cgi-bin/istor/journal.fcgi?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