Received: 19.8.2008 **Accepted: 8.10.2008**

Original Article

Usage of Penalized Maximum Likelihood Estimation Method in Medical Research: An Alternative to Maximum Likelihood Estimation Method

*Ecevit Eyduran**

Abstract

The paper was to reduce biased estimation using new approach (Penalized Maximum Likelihood Estimation (PMLE) Method) in Logistic Regression. For this aim, unreal four small data sets were randomly generated. Maximum Likelihood Estimation (MLE) and PMLE Methods were applied and compared for separation case including biased estimation in Logistic Regression when one of the cells in 2 x 2 tables becomes equal to zero (separation problem). Parameters$\square_1$ and their standard error obtained by using MLE for four data sets were 12.56 ± 257.8, 13.46 ± 264.3, 13.42±210.3, and 13.41 ± 180.4, respectively, meaning that MLE's are biased estimates. Corresponding values for PMLE method were found 2.28 ± 1.81, 3.05 ± 1.59, 3.45 ± 1.53, and 3.45 ± 1.53, respectively, meaning that PMLE's was unbiased estimates. It is clear that standard error value for data set 1 reduced from 257.8 to 1.81 when using PMLE method for separation problem. According to PMLE Method, the odds of being coronary heart disease risk for smokers were increased 21.08 times than that for non-smokers smoking in data set 2, which is significant at 1% level. The odds of being coronary heart disease risk for smokers were increased 31.63 times than that for non-smokers in data set 3 ($P < 0.001$). The odds of being coronary heart disease risk for smokers were increased 41.93 times than that for non-smokers in data set 4. When one of the cells in 2 x 2 contingency tables becomes equal to zero, PMLE was more superior to MLE Method because PMLE Method may be performed unbiased (reliable) estimation.

KEYWORDS: *Bias Shrinking, Penalized Maximum Likelihood Estimation, Logistic Regression.*

JRMS 2008; 13(6): 325-330

Chi-Square, Likelihood ratio Chi-Square and Logistic Regression have been used commonly in two-by-two (2x2) tables. It is well-known that Chi-Square and Likelihood ratio Chi-Square statistics are "goodness of fit" criteria in Logistic regression.[1-4] For analyzing medical studies with binary responses, Logistic Regression is frequently used. In logistic regression, the effect of explanatory variables on binary variable is explained by odds ratio. Odds ratio provides straightforward interpretation of estimated parameters. Parameter estimations in Logistic Regression are based on maximum likelihood method.[5-13]

The simplest form of logistic regression is a 2 x 2 table. When one of four cells in the two-by-two equal to zero, standard errors of parameters estimated by maximum likelihood method are too large and biased (unreliable). The phenomenon is known as separation or monotone likelihood. In separation case, converge operations on estimating parameters in SAS and SPSS statistical programs can not be performed. For example, SAS program gives some warnings: "*The maximum likelihood estimate may not exist*" and "*Validity of the model fit is questionable*".[5-12]

Some solutions to the separation problem have been suggested: arithmetic correction and profile penalized likelihood methods. Arithmetic corrections are based on adding ½ (the well-known Haldane Correction), 1 (the well-known Laplace Correction), and 2 Greenland et al. (2000) to one zero cell count in contingency tables and an advance method, an original approach called as profile penalized log

*University of Yüzüncü Yıl, Faculty of Agriculture, Department of Animal Science, Biometry Genetics Unit, 65080, Van-Turkey.
E-mail: ecevit.eyduran@gmail.com

likelihood (PPL) derived from firth's modified score test (FMCT). In the event of bias reduction of MLE's, FMCT suggested by Firth (1993) offers researchers finite parameter estimates via penalized maximum likelihood.[9,12]

Special macros for the advanced method were developed in package programs such as SAS, S-PLUS, and R.[10] It was reported that separation problem and biased estimations were eliminated by the special macros written in 3 package programs[12]. The three macros at different package program developed by Heinze and Ploner (2003) were based on profile penalized log likelihood (PPL) (which is also called Firth-type estimates) and Newton-Raphson algorithm was used for PPL parameter estimates in special SAS macro (at the web site:
http://www.meduniwien.ac.at/akh/imc/biometrie/programme/fl/). In sparse sample or separation problem, Penalized Likelihood Estimation removed O(n^{-1}) bias of Maximum Likelihood Estimation.[6-14]

It has been suggested in "Separation" case that Penalized Maximum Likelihood Estimation (Firth-Type Estimation) Method should be used instead of MLE in order to reduce biased estimation. The present paper aims to study how to reduce biased estimation when encountered Separation problem.

Methods

The four data sets were randomly generated to understand studied methods well. That is, unreal four data sets used in the present paper were arbitrarily generated.

Let's consider the link between coronary heart disease (CHD) risk and smoking (Table 1). The aim of this study is to evaluate the effect of smoking on coronary heart disease (CHD) risk using Maximum Likelihood and Penalized Maximum Likelihood Estimation Methods in small sample sizes. How many times the odds of high coronary heart disease (CHD) risk for smokers is higher than non-smokers?

Let's consider four various data sets on coronary heart disease (CHD) and smoking (Table 1). Consider that coronary heart disease (CHD) risk was assigned as dependent variable (with low and high levels) and smoking as explanatory variable (with yes and no levels). For example in data set 3, 17 patients who are smoking had low coronary heart disease (CHD) risk; 20 patients who are smoking had high coronary heart disease (CHD) risk; 18 patients who aren't smoking had low coronary heart disease (CHD) risk; 0 people who is not smoking had high coronary heart disease (CHD) risk. Risk factors of interest are high coronary heart disease (CHD) risk for dependent variable and smoking (yes) for explanatory variable.

The data were analyzed using a special SAS program (at the web site:
http://www.meduniwien.ac.at/msi/biometrie/programme/fl/).

P*enalized Maximum Likelihood Estimation (Firth Type Estimation)*

Logistic regression model is defined as $P(y_i = 1 \mid x_i) = \pi_i = 1/\{1+\exp(-x_i\beta)\}$ where $(y_i, x_i), y_i \in \{0,1\}, \quad i = 1,2,3\ldots,n$ denotes a sample of n observations of dependent variable *y* and the vector of independent variable with $(1xk)$ dimensions.

Maximum likelihood estimates for the regression parameters, namely, intercept and slopes β_r, $r = 1,2\ldots,k$ are found by solving the *k* score equations

$$\partial \log L / \partial \beta_r = U(\beta_r) = \sum_{i=1}^{n} (y_i - \pi_i) x_{ir} = 0,$$ where

L is the likelihood function. With intention of eliminating biased or infinite estimation of Maximum Likelihood in sparse samples and in the event of separation, Firth (1993) recommended to maximize $\log L(\beta)^* = \log L(\beta) + 0.5 \log |I(\beta)|$.

If the modification to a logistic model is applied

$$\Pr ob(y_i = 1 \mid x_i, \beta) = \pi_i = \left\{1 + \exp\left(-\sum_{r=1}^{k} x_{ir}\beta_r\right)\right\}^{-1}$$

then the score equation

$U(\beta_r) = \sum_{i=1}^{n}(y_i - \pi_i)x_{ir} = 0$ is replaced by the modified score equation

$$U(\beta_r)* = \sum_{i=1}^{n}\{y_i - \pi_i + h_i(1/2 - \pi_i)\}x_{ir} = 0$$

$(r = 1,...,k)$

Where, the h_i's are the i^{th} diagonal elements of the hat matrix $H = W^{1/2}X\left(X^TWX\right)^{-1}X^TW^{1/2}$, through $W = diag\{\pi_i(1-\pi_i)\}$. Firth-type (FL) estimates $\hat{\beta}$ can be obtained iteratively the usual way until convergence is obtained:

$$\beta^{(s+1)} = \beta^s + I^{-1}\left(\beta^{(s)}\right)U\left(\beta^{(s)}\right)^*$$

Where, the superscript (s) refers to the sth iteration.

Chi-Square and Likelihood Ratio Chi-Square Statistics

The notation of Chi-Square (1) and Likelihood Ratio Chi-Square statistics (2) are written as follows (Agresti, 2002):

$$\chi^2 = \sum\frac{(f - f_i)^2}{f_i} \qquad (1)$$

$$G = 2\sum f.\ln\left(\frac{f}{f_i}\right) \qquad (2)$$

Where, f is observed frequency and f_i is expected frequency.

Power Theory for Chi-Square and Likelihood Ratio Chi-Square Statistics

Assume that H_0 is the same to model M for a contingency table. Let π_i indicate the true probability in i^{th} cell and Let π_i(M) represent the value to which the Maximum likelihood (ML) estimate $\hat{\pi}_i$ for model M converges, where $\sum\pi_i = \sum\pi_i(M) = 1$. For multinomial sample of size n, the non-centrality parameter for Chi-Square (3) can be expressed as follows:

$$\lambda = n\sum_i\frac{[\pi_i - \pi_i(M)]^2}{\pi_i(M)} \qquad (3)$$

Expression 3 is the similar form as Chi-Square statistics, with for the sample proportion p_i and $\pi_i(M)$ in place of $\hat{\pi}_i$. The non-centrality parameter for Likelihood Ratio Chi-Square Statistics (4) can be written in this manner:

$$\lambda = 2n\sum_i\pi_i\log\frac{\pi_i}{\pi_i(M)} \qquad (4)$$

In order to obtain reliable results from both statistics, one should achieve a power value of at least 80%.[15] The highest value for power analysis of both statistics is 1. Power Analysis for Chi-Square and Likelihood Ratio Chi-Square Statistics were performed using a special SAS macro at web site: (http://ftp.sas.com/techsup/download/stat/powerrxc.html).

Results and Discussion

Maximum Likelihood and Penalized Likelihood Estimation

Table 2 presents Maximum Likelihood Estimations of parameters (β_0, β_1) after a logistic model were fitted to these four data sets in using PROC LOGISTIC of SAS package program. As table 2 shows, the values of parameters β_1 and their standard error for all data sets were found approximately:
12.56 ± 257.8, 13.46 ± 264.3, 13.42 ± 210.3, and 13.41 ± 180.4 respectively because one of four cells in 2 x 2 tables equals to zero (or separation case occurred). Odds ratio estimations and confidence interval for smoking was found as > 999.999 and (< 0.001, > 999.999) for the data sets. The parameter estimations are biased (undesired) estimations.

Programs written in statistical package programs such as in SAS, SPLUS, R were developed in order to solve separation problem. The programs are based on Penalized Maximum Likelihood, which was developed by Firth 1993.

To solve separation problem, Penalized Likelihood Estimation (profile penalized likelihood) were fitted to four data sets using a special SAS macro FL (http://www.meduniwien.ac.at/msi/biometrie/programme/fl/). Penalized Likelihood Estimations and standard errors of parameters are: 2.28 ± 1.81, 3.05 ± 1.59, 3.45 ± 1.53, and 3.45 ± 1.53, respectively as presented in Table 3. As Table 2 and 3 show, the usage of Penalized Likelihood Estimation method in separation problem reduced too large standard error values of parameters or biased. For example, Standard error value for data set 1 reduced from 257.8 to 1.81. The effect of smoking on coronary heart disease (CHD) in data set 1 was non-significant, whereas the effect of smoking on coronary heart disease (CHD) risk in other data sets were more significant ($P < 0.01$ for data set 2; $P < 0.001$ for data set 3 and 4).

Profile penalized likelihood Estimations via SAS macro FL written by Heinze and Ploner (2004) are presented in table 4. Odds ratio value was non-significant only for data set 1. Odds and interval estimates for smoking are presented in table 4.

The odds of coronary heart disease (CHD) for smokers were increased 21.08 times compared to non-smokers in data set 2, which is statistically significant at 1% level. The odds of coronary heart disease (CHD) for smokers were increased 31.63 times compared to non-smokers in data set 3 ($P < 0.001$). The odds of coronary heart disease (CHD) for smokers were increased 41.93 times compared to non-smokers in data set 4.

Chi-Square and Likelihood Ratio Chi-Square Statistics

All values on Chi-Square and Likelihood Ratio Chi-Square Statistics for each data set are presented in table 5. As table 5 shows, power values of these statistics calculated for data set 2, 3, and 4 were found to be 0.80 and more than 0.80 (very high level power values). However, corresponding values for data set 1 was power values with medium level (0.47 and 0.58). As sample size increased from 20 to 70, power values increased.

Required sample size (people number) to achieve 0. 80 power value in Chi-Square statistic were found as: 44 for data set 1, 35 for data set 2, 28 for data set 3 and 29 for data set 4 (data are not shown).

Required sample size (people number) to achieve 0. 80 power value in Likelihood Ratio Chi-Square statistic were found as 33 for data set 1, 26 for data set 2, 21 for data set 3, and 21 for data set 4 (data are not shown).

Table 1. The four sets of data at various sample sizes of Coronary Heart disease and Smoking.

Data Set 1	**Smoking (X)**	
Coronary Heart disease Risk (Y)	Yes (1)	No (0)
Low (0)	7	3
High (1)	10	0
Data Set 2	**Smoking (X)**	
Coronary Heart disease Risk (Y)	Yes (1)	No (0)
Low (0)	12	8
High (1)	15	0
Data Set 3	**Smoking (X)**	
Coronary Heart disease Risk (Y)	Yes (1)	No (0)
Low (0)	17	18
High (1)	20	0
Data Set 4	**Smoking (X)**	
Coronary Heart disease Risk (Y)	Yes (1)	No (0)
Low (0)	22	23
High (1)	25	0

Table 2. Maximum likelihood Estimates of parameters for four sets of data at various sample sizes on heart disease attack and smoking.

Parameters	Degrees of Freedom	Estimation of Parameters	Standard Error	Wald Statistics	Probability (P)	Odds Ratio
Data Set 1						
β_0	1	-0.3567	0.4928	0.5238	0.4692	-
β_1 (smoking)	1	12.5596	257.8	0.0024	0.9611	> 999.999
Data Set 2						
β_0	1	-0.2231	0.3873	0.3320	0.5645	-
β_1 (smoking)	1	13.4566	264.3	0.0026	0.9594	> 999.999
Data Set 3						
β_0	1	-0.1625	0.3299	0.2427	0.6223	-
β_1 (smoking)	1	13.4241	210.3	0.0041	0.9491	> 999.999
Data Set 4						
β_0	1	-0.1278	0.2923	0.1912	0.6619	-
β_1 (smoking)	1	13.4082	180.4	0.0055	0.9407	> 999.999

Table 3. Firth-Type estimates, profile penalized likelihood confidence limits.

Parameters	Degrees of Freedom	Estimation of Parameters	Standard Error	Lower 95% CL	Upper 95% CL	P > Chi-square
Data Set 1						
β_0	1	-0.33647	0.49195	-1.30070	0.62775	0.4785
β_1 (smoking)	1	2.28236	1.81372	-0.25207	7.24065	0.0816^{NS}
Data Set 2						
β_0	1	-0.21511	0.38713	-0.97388	0.54366	0.5704
β_1 (smoking)	1	3.04831	1.59128	0.80665	7.94638	0.0041**
Data Set 3						
β_0	1	-0.15822	0.32983	-0.80469	0.48824	0.6263
β_1 (smoking)	1	3.45406	1.53049	1.30986	8.33201	0.0003***
Data Set 4						
β_0	1	-0.15822	0.32983	-0.80469	0.48824	0.6263
β_1 (smoking)	1	3.45406	1.53049	1.30986	8.33201	0.0003***

NOTE: Confidence interval for Intercept based on Wald method.
NS: Non-Significant *: (P < 0.05) **: (P < 0.01) ***: (P < 0.001)

Table 4. FL odds ratio estimates, profile penalized likelihood confidence limits.

Effect	Odds Ratio	Pr > Chi Sq
Data set 1 (smoking)	9.79974	0.0816^{NS}
Data set 2 (smoking)	21.0797	0.0041**
Data set 3 (smoking)	31.6285	0.0003***
Data set 4 (smoking)	41.9333	<.0001***

NS: Non-Significant *: (P < 0.05) **: (P < 0.01) ***: (P < 0.001)

Table 5. Goodness of fit values and their power values for each data set.

	N	Chi-Square	L.R. Chi-Square	Power Value of Chi-Square	Power Value of L.R. Chi-Square
Data Set 1	20	3.53ns	4.69*	0.470	0.580
Data Set 2	35	7.78**	10.71**	0.800	0.910
Data Set 3	55	15.29***	21.05***	0.970	0.996
Data Set 4	70	19.03***	26.28***	0.992	0.999

Conclusion

In all scientific areas, parameter estimates in Logistic regression are biased for Maximum Likelihood Estimation when one of the cells in 2 x 2 contingency tables equal to zero. With the usage of penalized maximum likelihood approach developed by Firth (1993), biased estimates due to separation problem are reduced. These finding were generally consistent with those reported earlier on separation problem.

To obtain unbiased estimation in logistic regression:

1. Total sample size should be increased.
2. Many explanatory variables (such as male sex, increasing age, heredity (including race), high blood cholesterol, high blood pressure, physical inactivity, obesity, and diabetes mellitus) should be added to logistic regression model.

As a result, it was concluded that in separation problem, Penalized Maximum Likelihood estimation Method was superior to Maximum Likelihood Estimation Method.

References

1. Allison PD. Logistic regression using the SAS system: theory and application. 1st ed. Cary, NC: SAS Publishing, 1999.
2. Hosmer DW, Lemeshow S. Applied logistic regression. 2nd ed. New York: Wiley-Interscience Publication, 2000.
3. Sharma S .Applied multivariate techniques. New York: John Wiley & Sons, 1996.
4. Eyduran E, Özdemir T, Çak B, Alarslan E. Using of logistic regression in Animal Science. Ansinet, Journal of Applied Sciences 2005; 5(10):1753-6.
5. Eyduran E, Ozdemir T. Unbiased estimation for logistic regression. Journal of Science and Technology 2007; 1(2):245-251.
6. Firth D. Bias reduction, the Jeffreys prior and GLIM. In: Fahrmeir L, Francis B, Gilchrist R, Tutz G, editors. Advances in GLIM and statistical modelling. New York: Springer-Verlag, 1992; 91-100.
7. Firth D. Generalized linear models and Jeffreys priors: an iterative weighted least- squares approach. In: Dodge Y, Whittaker J, editors. Computational statistics. Heidelberg: Physica-Verlag, 1992: 553-557.
8. Firth D. Bias reduction of maximum likelihood estimates. Biometrika 1993; 80(1):27-38.
9. Bull SB, Mak C, Greenwood C.M.T. A modified score function estimator for multinomial logistic regression in small samples. Computational Statistics and Data Analysis 2002; 39:57-74.
10. Heinze G, Schemper M. A solution to the problem of separation in logistic regression. Stat Med 2002; 21(16):2409-2419.
11. Heinze G, Ploner M. Fixing the nonconvergence bug in logistic regression with SPLUS and SAS. Computer Methods and Programs in Biomedicine 2003; 71:181-187.
12. Heinze G, Ploner M. A SAS macro, S-PLUS library and R package to perform logistic regression without convergence problems. 2. 2004. Austria, Section of Clinical Biometrics, Department of Medical Computer Sciences , Medical University of Vienna
13. (http://www.meduniwien.ac.at/akh/imc/biometrie/programme/fl/)
14. Greenland S. Small-sample bias and corrections for conditional maximum-likelihood odds-ratio estimators. Biostatistics 2000; 1(1):113-122.
15. Agresti A. Categorical data analysis. 2nd ed. New Jersey: Wiley-Interscience, 2002.