PAPER • OPEN ACCESS

The parameter estimation of logistic regression with maximum likelihood method and score function modification

To cite this article: R Febrianti *et al* 2021 *J. Phys.: Conf. Ser.* **1725** 012014

The parameter estimation of logistic regression with maximum likelihood method and score function modification

R Febrianti, Y Widyaningsih and S Soemartojo

Department of Mathematics, Faculty of Mathematics and Natural Sciences (FMIPA), Universitas Indonesia, Depok 16424, Indonesia

Corresponding author's email: yekti@sci.ui.ac.id

Abstract. The maximum likelihood parameter estimation method with Newton Raphson iteration is used in general to estimate the parameters of the logistic regression model. Parameter estimation using the maximum likelihood method cannot be used if the sample size and proportion of successful events are small, since the iteration process will not yield a convergent result. Therefore, the maximum likelihood method cannot be used to estimate the parameters. One way to resolve this un-convergence problem is using the score function modification. This modification is used to obtain the parameters estimate of logistic regression model. An example of parameter estimation, using maximum likelihood method with small sample size and proportion of successful events equals 0.1, showed that the iteration process is not convergent. This non-convergence can be solved with modifications on a score function. Modification on score function is to change a score function, a matrix of the first derivative of the log likelihood function, to the first derivative matrix itself minus multiplication of information matrix and biased vector. The modification of the score function can quickly yield values of parameter estimates, especially when the sample sizes are larger, and convergence was reached before the 10th iteration.

Keywords: Maximum likelihood, score function modification

1. Introduction

Regression analysis is a statistical method to analyze the relationship between one response variable and one or more explanatory variables. Regression analysis is used to analyze data with quantitative response variable. If the response variable is a qualitative variable, the linear regression model cannot be used. So, logistic regression model is used to analyze the data with qualitative response variable. Logistic regression model is defined as [1].

$$\pi(x_i) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_i)}}$$

or

$$\pi(x_i) = \frac{e^{(\beta_0 + \beta_1 x_i)}}{1 + e^{(\beta_0 + \beta_1 x_i)}}$$

where $i = 1, 2, .., n$.

The maximum likelihood method can be used to estimate the parameters of the logistic regression model. Y_i is a random variable and is independent with $i = 1,2,\dots,n$ and $Y_i \sim Bernoulli\ (\pi(x_i)$. Defined likelihood function of $Y_i \sim Bernoulli\ (\pi(x_i))$ is [2, 3].

$$p(y_1, y_2, \dots y_n) = \prod_{i=1}^{n} \pi(x_i)^{y_i}\left(1 - \pi(x_i)\right)^{1-y_i} \tag{1}$$

Because $\frac{\pi(x_i)}{1-\pi(x_i)} = e^{(\beta_0 + \beta_1 x_i)}$ and $1 - \pi(x_i) = \frac{1}{1+e^{(\beta_0 + \beta_1 x_i)}}$ then equation 1 can be expressed as:

$$p(y_1, \dots y_n, \beta_0, \beta_1) = L(\beta_0, \beta_1) = \prod_{i=1}^{n} (e^{(\beta_0 + \beta_1 x_i)})^{y_i} \frac{1}{1 + e^{(\beta_0 + \beta_1 x_i)}} \tag{2}$$

To simplify the derivative of likelihood function, the log-likelihood function is used. The log-likelihood function is defined as follows:

$$l(\beta_0, \beta_1) = \ln(L(\beta_0, \beta_1)) = \sum_{i=1}^{n} y_i(\beta_0 + \beta_1 x_i) - \sum_{i=1}^{n} \ln\ \left(1 + e^{\beta_0 + \beta_1 x_i}\right) \tag{3}$$

Next, the derivative of log-likelihood function with respect to β_0, is [4].

$$\frac{\partial l(\beta_0, \beta_1)}{\partial \beta_0} = \sum_{i=1}^{n} y_i - \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \tag{4}$$

Since $\pi(x_i) = \frac{e^{\beta_0 + \beta_1 x_i}}{1+e^{\beta_0 + \beta_1 x_i}}$ then equation 4 can be expressed as:

$$\frac{\partial l(\beta_0, \beta_1)}{\partial \beta_0} = \sum_{i=1}^{n} (y_i - \pi(x_i)) = 0 \tag{5}$$

The derivative of log-likelihood function with respect to β_1 is

$$\frac{\partial l(\beta_0, \beta_1)}{\partial \beta_1} = \sum_{i=1}^{n} y_i x_i - x_i \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \tag{6}$$

Again, since $\pi(x_i) = \frac{e^{\beta_0 + \beta_1 x_i}}{1+e^{\beta_0 + \beta_1 x_i}}$, equation 6 can be expressed as:

$$\frac{\partial l(\beta_0, \beta_1)}{\partial \beta_1} = \sum_{i=1}^{n} (y_i x_i - x_i \pi(x_i)) = \sum_{i=1}^{n} (y_i - \pi(x_i))x_i = 0 \tag{7}$$

Equation 5 and equation 7 are not in β_0 and β_1, and are difficult to find the solutions analytically, then to obtain the values of $\hat{\beta}_0$ and $\hat{\beta}_1$, Newton Raphson numerical iteration method should be used [5].

The steps to estimate parameter $\boldsymbol{\beta}$ using Newton-Raphson iteration are

1. Input the initial estimated value of $\boldsymbol{\beta}_{(\mathbf{0})}$
2. To obtain estimation values on the *(k+1)-th* iteration, calculate $\boldsymbol{\beta}_{(k+1)} = \boldsymbol{\beta}_{(k)} - \boldsymbol{I}(\boldsymbol{\beta})^{-1}{}_{(k)}\boldsymbol{U}(\boldsymbol{\beta})_{(k)}$.
3. The iteration is continued until $\boldsymbol{\beta}_{(k+1)} \approx \boldsymbol{\beta}_{(k)}$.

$\boldsymbol{U}(\boldsymbol{\beta})$ is defined as a matrix of the first derivative of log likelihood function with respect to the parameters

$$\boldsymbol{U}(\boldsymbol{\beta}) = \begin{bmatrix} \dfrac{\partial l(\boldsymbol{\beta})}{\partial \beta_0} \\ \dfrac{\partial l(\boldsymbol{\beta})}{\partial \beta_1} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} y_i - \pi(x_i) \\ \sum_{i=1}^{n} y_i x_i - x_i\,\pi(x_i) \end{bmatrix}$$

and $\boldsymbol{I}(\boldsymbol{\beta})^{-1}$ is

$$\boldsymbol{I}(\boldsymbol{\beta}) = \begin{bmatrix} \dfrac{\partial^2 l(\beta_0, \beta_1)}{\partial \beta_0^2} & \dfrac{\partial^2 l(\beta_0, \beta_1)}{\partial \beta_0 \partial \beta_1} \\ \dfrac{\partial^2 l(\beta_0, \beta_1)}{\partial \beta_1^2} & \dfrac{\partial^2 l(\beta_0, \beta_1)}{\partial \beta_1 \partial \beta_0} \end{bmatrix}^{-1}$$

$$\boldsymbol{I}(\boldsymbol{\beta}) = -\begin{bmatrix} \sum_{i=1}^{n} \pi(x_i)\big(1-\pi(x_i)\big) & \sum_{i=1}^{n} x_i\pi(x_i)\big(1-\pi(x_i)\big) \\ \sum_{i=1}^{n} x_i\pi(x_i)\big(1-\pi(x_i)\big) & \sum_{i=1}^{n} x_i^2\pi(x_i)\big(1-\pi(x_i)\big) \end{bmatrix}^{-1}$$

2. Estimation of parameter using modification of score function

The logistic regression model uses the maximum likelihood method to estimate the parameters of the model, using Newton Raphson method to obtain the final solution. According to Badi N H S [6], the Newton Raphson iteration is not convergent if the sample size is small and the proportion of success events is small. According to Czepiel S A [7] if the result of parameter estimation through the iteration is not convergent, indicate that the model formed is no suitable for the data being analyzed.

The solution to solve the divergence problem in Newton Raphson iteration is to modify the score function. Modification on score function discovered by Firth in 1993. Modification of score function is using bias vector and information matrix to estimate parameter in logistic regression model. Mathematically, the modification of score functions is to change $\boldsymbol{U}(\boldsymbol{\beta})$ to $\boldsymbol{U}^*(\boldsymbol{\beta})$ as follows [8]:

$$\boldsymbol{U}^*(\boldsymbol{\beta}) = \boldsymbol{U}(\boldsymbol{\beta}) - \boldsymbol{\mathcal{I}}(\boldsymbol{\beta})\boldsymbol{b}(\boldsymbol{\beta}) = 0 \tag{8}$$

where $\boldsymbol{\mathcal{I}}(\boldsymbol{\beta})$ is an information matrix, defined as:

$$\boldsymbol{\mathcal{I}}(\boldsymbol{\beta}) = \begin{bmatrix} -E\left(\dfrac{\partial^2 l(\beta_0, \beta_1)}{\partial \beta_0 \partial \beta_0}\right) & -E\left(\dfrac{\partial^2 l(\beta_0, \beta_1)}{\partial \beta_0 \partial \beta_1}\right) \\ -E\left(\dfrac{\partial^2 l(\beta_0, \beta_1)}{\partial \beta_1 \partial \beta_0}\right) & -E\left(\dfrac{\partial^2 l(\beta_0, \beta_1)}{\partial \beta_1 \partial \beta_1}\right) \end{bmatrix}$$

$$\mathcal{I}(\boldsymbol{\beta}) = \begin{bmatrix} \sum_{i=1}^{n} \pi(x_i)(1-\pi(x_i)) & \sum_{i=1}^{n} x_i\,\pi(x_i)\,(1-x_i\,\pi(x_i)) \\ \sum_{i=1}^{n} x_i\,\pi(x_i)(1-\pi(x_i)) & \sum_{i=1}^{n} x_i^2\,\pi(x_i)(1-\pi(x_i)) \end{bmatrix} = \boldsymbol{X^TWX}$$

with $\boldsymbol{W} = diag\ \{\pi(x_i)(1-\pi(x_i))\}$ and $\boldsymbol{X}$ is the *design matrix.*
$\boldsymbol{b}(\boldsymbol{\beta})$ is a bias vector, defined as:

$$\boldsymbol{b}(\boldsymbol{\beta}) = \left(\boldsymbol{X^TWX}\right)^{-1}\boldsymbol{X^TW}\,\boldsymbol{\xi}$$

where

$$\boldsymbol{\xi} = \begin{bmatrix} \frac{h_{11}}{\pi(x_1)(1-\pi(x_1))}\left(\pi(x_1)-\frac{1}{2}\right) \\ \frac{h_{22}}{\pi(x_2)(1-\pi(x_2))}\left(\pi(x_2)-\frac{1}{2}\right) \\ \vdots \\ \frac{h_{nn}}{\pi(x_n)(1-\pi(x_n))}\left(\pi(x_n)-\frac{1}{2}\right) \end{bmatrix}$$

with h_{ii} being an element of the diagonal of the *hat* matrix. In *Generalized Linear Models*, the *hat* matrix was calculated as:

$$\boldsymbol{H} = \boldsymbol{W^{\frac{1}{2}}X}\left(\boldsymbol{X^TWX}\right)^{\boldsymbol{-1}}\boldsymbol{X^TW^{\frac{1}{2}}}$$

So the formula of modification of score function of regression logistics model is:

$$\begin{aligned} \boldsymbol{U}^*(\boldsymbol{\beta}) &= \boldsymbol{U}(\boldsymbol{\beta}) - \boldsymbol{\mathcal{I}}(\boldsymbol{\beta})\boldsymbol{b}(\boldsymbol{\beta}) = 0 \\ 0 &= \boldsymbol{U}(\boldsymbol{\beta}) - \left(\boldsymbol{X^TWX}\right)\left(\boldsymbol{X^TWX}\right)^{-1}\boldsymbol{X^TW}\,\boldsymbol{\xi} \\ 0 &= \boldsymbol{U}(\boldsymbol{\beta}) - \boldsymbol{IX^TW}\,\boldsymbol{\xi} \\ 0 &= \boldsymbol{U}(\boldsymbol{\beta}) - \boldsymbol{X^TW}\,\boldsymbol{\xi} \\ \begin{bmatrix} 0 \\ 0 \end{bmatrix} &= \begin{bmatrix} \sum_{i=1}^{n}[(y_i + \frac{h_{ii}}{2}) - (1+h_{ii})\pi(x_i)] \\ \sum_{i=1}^{n}[(y_i + \frac{h_{ii}}{2}) - (1+h_{ii})\pi(x_i)]\,x_i \end{bmatrix} = \begin{bmatrix} \boldsymbol{U_0^*} \\ \boldsymbol{U_1^*} \end{bmatrix} \end{aligned}$$

The result of $\boldsymbol{\beta}$ parameter estimation for modification of score functions requires numeric iteration. The *(m+1)*-th iteration of modification of score function is:

$$\boldsymbol{\beta}_{m+1}^* = \boldsymbol{\beta}_m^* - \boldsymbol{b}(\boldsymbol{\beta}_m^*) + \boldsymbol{I}_{(m)}(\boldsymbol{\beta})^{-1}\,\boldsymbol{U}^*(\boldsymbol{\beta}_m^*)$$

with $\boldsymbol{\beta}_m^*$ is the value of $\boldsymbol{\beta}$ at the *m*-th iteration,
$\boldsymbol{b}(\boldsymbol{\beta}_m^*)$ is bias vector at the *m*-th iteration,
$\boldsymbol{I}_{(m)}(\boldsymbol{\beta})^{-1}$ is the invers of *information matrix* at the *m*-th iteration,
$\boldsymbol{U}^*(\boldsymbol{\beta}_m^*)$ is the score function at the *m*-th iteration,
$m \geq 0$.

3. Application

The maximum likelihood parameter estimation and modification of score function to logistic regression models is applied on endometrial cancer data. In this data, HG (Histology Grade) is a high or low value of endometrial cancer that is determined as variable response. If the HG value is 1, cancer endometrial is on high stadium; if HG value is 0, cancer endometrial is on low stadium. EH (Endometial Hyperplasia) is state of the endometrial growing to excess. EH is the explanatory variable in modeling. The data consists of 79 observations, with response values of y = 1 are 30 observations and the response values of y = 0 are 49 observations [7]. The samples application use samples of size n = 10, n = 20, and n = 30 with the proportion of y = 1 is 0.1 and the stopping criterion in the program is the maximum iteration of 10,000 or the error tolerance in the program is 1×10^{-6}.

Table 1 shows that the parameter estimation using maximum likelihood method with Newton Raphson iteration is not convergent. This problem is solved using modification of the score function to estimate the parameters of the model. Table 2 is the result of a modification of score function for n = 10 and proportion of y = 1 is 0.1.

Table 2 shows that the result of the parameter estimation using a modification of the score function gives the values of $\hat{\beta}_0$ = -5.7266 and $\hat{\beta}_1$= 1.9464. A modification on the score function is able to solve un-convergence parameter estimation problem of maximum likelihood using Newton Raphson iteration. Convergence begins at the 564th iteration.

Table 1. The results of maximum likelihood parameter estimation using *Newton Raphson* iteration without modification of score function, n = 10, proportion of y = 1 is 0.1.

Iteration	$\hat{\beta}_0$	$\hat{\beta}_1$
1	-4.9355	2.3556
2	-8.3121	4.0197
3	-11.5656	5.6104
4	-14.8322	7.1973
5	-21.579	8.8085
⋮	⋮	⋮
⋮	⋮	⋮
10,000	$-\infty$	$-\infty$

Table 2. The results of maximum likelihood parameter estimation using *Newton Raphson* iteration with modification of the score function, n = 10, proportion of y = 1 is 0.1

Iteration	$\hat{\beta}_0$	$\hat{\beta}_1$
1	-4.9355	2.3556
2	-4.9355	2.2556
3	-4.9355	2.1556
4	-5.0355	2.1556
5	-5.1355	2.1556
⋮	⋮	⋮
563	-5.7267	1.9463
564	-5.7266	1.9464

Table 3. The results of maximum likelihood parameter estimation using *Newton Raphson* iteration with modification of score function, n = 20, proportion of y = 1 is 0.1.

Iteration	$\hat{\beta}_0$	$\hat{\beta}_1$
1	0.1741	-0.9404
2	2.5349	-2.6134
3	6.0205	-5.0421
4	9.1708	-7.4126
5	11.8665	-9.4988
6	12.6815	-10.1196
7	13.0317	-10.3947
8	13.7242	-10.9536
9	13.7242	-10.9536
10	13.7242	-10.9536

Table 4. The results of maximum likelihood parameter estimation using *Newton Raphson* iteration with modification of score function, n = 30, proportion of y = 1 is 0.1.

Iteration	$\hat{\beta}_0$	$\hat{\beta}_1$
1	-0.1838	-0.7757
2	1.2144	-1.9532
3	2.9857	-3.3758
4	4.1439	-4.3899
5	4.5526	-4.7656
6	4.5949	-4.8054
7	4.5949	-4.8054
8	4.5949	-4.8054

Table 3 and table 4 are the results of iteration for sample sizes of 20 and 30, respectively, with a proportion of success of 0.1 using the maximum likelihood parameter estimation method using *Newton Raphson* iteration with modification of the score function.

Based on the results of tables 3 and table 4, the maximum likelihood estimation method with modification on the score function using Newton Raphson iteration with larger sample sizes can give values of parameter estimation rapidly. For sample size of n = 20, the convergence parameter starts on the 8^{th} iteration. Furthermore, for sample size of n = 30, the convergence parameter starts on the 6^{th} iteration.

4. Conclusion

To estimate the parameters of the logistic regression model using the maximum likelihood method is to differentiate the likelihood function, then set this first derivative to 0, and continue to solve the equation to obtain the estimate of parameters. The first derivative of the likelihood function on the parameters is not linear and it is difficult to obtain the solution analytically. Therefore, it required Newton-Raphson iteration. Here, the iterations never gave a convergent result. Furthermore, modification on score functions is needed, that is, using bias vectors and information matrices to estimate parameters in the logistic regression model. Mathematically, the purpose of modification of score functions is to change a score function that is the first derivative matrix, to the first derivative matrix itself minus multiplication of information matrix and biased vectors. The modification of score functions can quickly yield values of parameter estimation. Based on the results of computations to estimate parameters, if the sample size is small and the proportion of success events is also small, using the maximum likelihood method with Newton-Raphson iteration will not working properly. This problem can be solved using modification of score functions.

References

[1] Agresti A 2015 *Foundations of Linear and Generalized Linear Models* (Hoboken: Wiley)
[2] Hogg R V and Craig A 1995 *Introduction To Mathematical Statistics* 5th edition (New Jersey: Prentice Hall)
[3] Hosmer D W and Lemeshow S 2000 *Applied Logistic Regression* 2nd edition (Hoboken: Wiley)
[4] Montgomery D C, Peck E A and Vinning G G *2001 Introduction to Linear Regression Analysis* 3rd edition (Hoboken: John Wiley & Sons, Inc)
[5] Pawitan Y 2001 *In All Likelihood: Statistical Modelling and Inference Using Likelihood* (Oxford: Oxford University Press)
[6] Badi N H S 2017 *Open Access Lib. J.* **4** e3625
[7] Czepiel S A 2003 *Maximum likelihood Estimation of Logistic Regression Models: Theory and Implementation* available at https://czep.net/stat/mlelr.pdf
[8] Firth D 1993 *Biometrika* **80** 27-38