

Modified maximum likelihood estimation in poisson regression

Abstract

In Generalized Linear Models, likelihood equations are intractable and do not have explicit solutions; thus, they must be solved by using Newton-type algorithms. Solving these equations by iterations, however, can be problematic: the iterations might converge to wrong values or the iterations might not converge at all. In this study, we derive the modified maximum likelihood estimators for Poisson regression model and study their properties. We also search the robustness of these estimators when there are outliers in the covariates.

Keywords: count data, poisson regression, modified maximum likelihood, newton-type algorithms, dixon's outlier model

Volume 6 Issue 1 - 2017

Evrin Oral

Department of Biostatistics Program, LSUHSC School of Public Health, USA

Correspondence: Evrin Oral, Biostatistics Program, LSUHSC School of Public Health, 2020 Gravier Street, 3rd floor New Orleans, LA 70112, USA, Tel (504) 568-6094, Email eoral@lsuhsc.edu

Received: April 25, 2017 | **Published:** May 25, 2017

Introduction

Poisson regression is widely used for modeling count data, especially when there is no over- or under- dispersion.¹ Since the likelihood equations from this model are intractable, solving these equations requires using iterative methods, such as Newton Raphson or Fisher scoring. However, using iterative methods to find maximum likelihood estimators (MLEs) can generally be problematic and time should be spent to investigate the stability of such solutions.^{2,4} Specifically the following difficulties can arise: the iterations might converge to wrong values if the likelihood equations have multiple roots, or the iterations might not converge at all. See^{5,6} for a discussion about situations where one encounters these difficulties in solving MLEs. The common software, Stata for example, is known to be very sensitive to numerical iterations. Researchers have reported problems in getting Poisson regression estimates with the "poisson" command, which encounters problems in locating the maximum and does not converge.⁷ Note that in Poisson regression modeling, additional problems might occur. The most common problem is the over- or under- dispersion in data, in which case using a more flexible model such as negative binomial regression is more appropriate. The second problem is analogous to the complete separation or quasi-complete separation problem in binary regression: the MLEs may not exist for certain data configurations, see.^{7,8} In this study we do not consider either of these problems and focus only on the case where it is appropriate to model the data with Poisson regression.

Unlike the maximum likelihood (ML) technique, modified maximum likelihood (MML) methodology produces explicit estimates. MML achieves explicit estimates by linearizing intractable functions within the likelihood equations using the ordered statistics.⁹ Asymptotically, MML estimators (MMLEs) are known to be unbiased and have minimum variances, i.e. they are fully efficient. For small sample sizes MMLEs have negligible bias and their variances are only marginally bigger than the minimum variance bounds, i.e. they are highly efficient.¹⁰⁻¹⁵

In the GLM setting, Tiku and Vaughan³ used the MML methodology to extend the techniques of traditional logistic

regression to non-logistic density functions. Oral and Gunay¹⁶ and Oral¹⁷ later extended the work in³ to the binary regression model with one stochastic covariate. Oral¹⁸ derived the MMLEs in general GLMs which use canonical link when there is only one risk factor. In this study, we derive the explicit MMLEs of Poisson regression model, generalize the derivations to more than one covariate, and study their robustness properties via simulations.

Methods

The Univariate Poisson regression model is given by

$$E(Y_i | X_i = x_i) = \mu_i \\ = \exp(z_i) \quad (1)$$

where $z_i = \alpha + \beta x_i$, for $1 \leq i \leq n$, and the outcome Y has the probability distribution

$$f_y(y_i) = \frac{\exp(-\mu_i) \mu_i^{y_i}}{y_i!}, y_i = 0, 1, 2, \dots \quad (2)$$

Note that in equation (2), Y is presumed to increase with X so that β is a priori greater than zero. For the model given in (1)-(2), the log-likelihood function of the random sample (y_i, x_i) can be written as

$$\ln L \propto \sum_{i=1}^n y_i z_i - \sum_{i=1}^n g(z_i), \quad (3)$$

where $g(z_i) = \exp(z_i)$. The likelihood equations for estimating α and β do not have explicit solutions because of the nonlinear function $g(z_i)$, $1 \leq i \leq n$. To obtain the MMLEs, we first express the likelihood equations in terms of the ordered variates $z_{(1)} \leq z_{(2)} \leq \dots \leq z_{(n)}$. The likelihood equations can be re-written as

$$\frac{\partial \ln L}{\partial \alpha} = \sum_{i=1}^n \{y_{[i]} - g(z_{(i)})\} = 0 \quad (4)$$

and

$$\frac{\partial \ln L}{\partial \beta} = \sum_{i=1}^n x_{(i)} \{y_{(i)} - g(z_{(i)})\} = 0 \quad (5)$$

Where $x_{(i)}$ is the concomitant of $x_{(i)}$ ($1 \leq i \leq n$) and $g(z_{(i)}) = \exp(z_{(i)})$. Linearizing the intractable function $g(z_i)$ by using the first two terms of its Taylor series expansion around the population quantiles $t_{(i)} = E(z_{(i)})$ ($1 \leq i \leq n$), we find

$$g(z_{(i)}) \approx a_i + b_i z_{(i)}, \quad (6)$$

where $a_i = \exp(t_{(i)}) \{1 - t_{(i)}\}$ and $b_i = \exp(t_{(i)})$ for $1 \leq i \leq n$. In order to calculate $t_{(i)} = E(z_{(i)})$ values, we define a dummy random variable U with the probability density function

$$f(u) = \exp(u), \quad u < 0 \quad (7)$$

and re-write the model (1) as

$$E(Y_i | X_i = x_i) = \mu_i = F(z_i), \quad (8)$$

where $F(u) = \exp(u)$, $u < 0$. Thus, the $t_{(i)}$ values can be obtained from the equation

$$t_{(i)} = \ln(i/(n+1)), \quad (9)$$

For $1 \leq i \leq n$; asymptotically $z_{(i)} - t_{(i)} \approx 0$. Alternatively, when n is large one can utilize the standard normal distribution

$$t_{(i)} = \Phi^{-1}(i/(n+1)). \quad (10)$$

Incorporating (6) into (4)-(5) and solving the resulting modified likelihood equations yield the explicit MMLEs below:

$$\hat{\alpha} = \frac{\delta}{m} - \hat{\beta} \bar{x}_a \text{ and } \hat{\beta} = \frac{\sum_{i=1}^n \delta_i (x_{(i)} - \bar{x}_a)}{\sum_{i=1}^n b_i (x_{(i)} - \bar{x}_a)^2}, \quad (10)$$

where

$$\delta = \sum_{i=1}^n \delta_i, \quad \delta_i = y_{(i)} - a_i, \quad m = \sum_{i=1}^n b_i, \quad (11)$$

and

$$\bar{x}_a = \left(\sum_{i=1}^n b_i x_{(i)} \right) / m. \quad (12)$$

The MMLEs derived above are asymptotically equivalent to their corresponding MLEs, giving them the same attractive asymptotic properties; however, one can refine the estimates by re-calculating a_i and b_i values by replacing the theoretical population quantiles with their estimated values $t_{(i)} = \hat{\alpha} + \hat{\beta} x_{(i)}$ ($1 \leq i \leq n$). This process might be repeated until a desired convergence is met. The stabilization generally is reached within a few iterations.

Asymptotic variances and co-variances

Vaughan and Tiku¹⁵ proved rigorously that the MMLEs are asymptotically unbiased and their variances and co-variances are exactly the same as those of the MLEs. In the present situation, therefore, the asymptotic variances and the covariance of $\hat{\alpha}$ and $\hat{\beta}$ are given by $I^{-1}(\alpha, \beta)$, where I is the Fisher Information matrix consisting of the elements $-E(\partial^2 \ln L^* / \partial \alpha^2)$, $-E(\partial^2 \ln L^* / \partial \alpha \partial \beta)$, and $-E(\partial^2 \ln L^* / \partial \beta^2)$. From the modified likelihood equations

$$\frac{\partial \ln L}{\partial \alpha} - \frac{\partial \ln L^*}{\partial \alpha} = \sum_{i=1}^n \{y_i - (a_i + b_i z_i)\} = 0,$$

$$\frac{\partial \ln L}{\partial \beta} - \frac{\partial \ln L^*}{\partial \beta} = \sum_{i=1}^n x_i \{y_i - (a_i + b_i z_i)\} = 0,$$

the Fisher Information matrix can be easily obtained as

$$V = I^{-1}(\gamma_0, \gamma_1) = \begin{bmatrix} \sum_{i=1}^n Q_i & \sum_{i=1}^n Q_i x_i \\ \sum_{i=1}^n Q_i x_i & \sum_{i=1}^n Q_i x_i^2 \end{bmatrix}^{-1}, \quad (13)$$

where $Q_i = \exp(z_i)$. V is estimated by replacing Q_i with its estimate $\hat{Q}_i = \exp(\hat{z}_i)$, $\hat{z}_i = \hat{\alpha} + \hat{\beta} x_i$ ($1 \leq i \leq n$). Since z_i values converge to t_i values as n tends to infinity, a_i and b_i values are treated as constant coefficients for large n , see also.³ Hence, the asymptotic variances can be estimated by

$$\text{Var}(\hat{\alpha}) = \sum_{i=1}^n \hat{Q}_i x_i^2 / \left\{ \sum_{i=1}^n \hat{Q}_i \sum_{i=1}^n \hat{Q}_i x_i^2 - \left(\sum_{i=1}^n \hat{Q}_i x_i \right)^2 \right\}, \quad (14)$$

$$\text{Var}(\hat{\beta}) = \sum_{i=1}^n \hat{Q}_i / \left\{ \sum_{i=1}^n \hat{Q}_i \sum_{i=1}^n \hat{Q}_i x_i^2 - \left(\sum_{i=1}^n \hat{Q}_i x_i \right)^2 \right\}. \quad (15)$$

Hypothesis testing

Testing the null hypothesis $H_0: \beta = 0$ is of great practical importance in Poisson regression modelling. The likelihood ratio statistic for testing H_0 is $LR = -2(L_0 - L_1)$, where L_0 and L_1 denote the maximized log-modified likelihood functions under the null and alternative hypotheses, respectively. The null distribution of LR is asymptotically a chi-square with 1 degree of freedom. Large values of LR lead the rejection of H_0 . Alternatively, the Wald statistic W (the ratio of $\hat{\beta}$ to its standard error) might be used. Since $\hat{\beta}$ is asymptotically equivalent to the MLE, the null distribution of W is asymptotically normal $N(0, 1)$. Large values of W lead to the rejection of H_0 .

Numerical example

To compare ML and MML estimates numerically, we analyzed the data given on page 82 of Agresti.¹⁹ The data is from a study of nesting horseshoe crabs where the response Y is the number of satellites that each female crab has, and the corresponding values of the covariate

X is the carapace width of 173 crabs. The study investigates the relationship between Y and X . We calculated the MMLs from equations (10)-(12) and their approximate standard errors from (14)-(15). The FORTRAN code written to carry out the calculations can be obtained from the author. Our results are completely consistent with those given in,¹⁹ which is expected; see Table 1.

Table 1 MLEs and MMLs along with their standard errors for horseshoe crab data

	Coefficient	Estimate	SE	W	LR
ML	α	-3.3048	0.5422		
	β	0.164	0.02	8.2	64.9
	Coefficient	Estimate	SE	W	LR
MML	α	-3.3047	0.5423		
	β	0.164	0.0199	8.241	64.91

Remark: In solving (10)-(12), Oral¹⁸ proposed to calculate the initial values from the least squares estimators (LSEs), which is a different approach than using equation (9) (Approach 1) or equation (10) (Approach 2). Since $t_{(i)} = E(\alpha + \beta x_{(i)})$, $t_{(i)}$ values can be approximated by $\tilde{t}_{(i)} = \tilde{\alpha} + \tilde{\beta} x_{(i)}$, where

$$\tilde{\alpha} = \bar{y} - \tilde{\beta} \bar{x} \text{ and } \tilde{\beta} = \frac{\sum_{i=1}^n (x_i - \bar{x}) y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

are the LSEs; see also.^{3,20} When using this approach (say, Approach 3), the $t_{(i)}$ values need to be revised after the first iteration with their estimated values $t_{(i)} = \hat{\alpha} + \hat{\beta} x_{(i)}$ ($1 \leq i \leq n$) as described above. Estimating population quantiles from the LSEs changes neither the derivations/solutions (10)-(12) nor the results. However, the total revision number needed for stabilization under different approaches is not the same, see also³ page 889. To compare the performance of these three approaches, we conducted a simulation study where we calculated the bias values and variances of the resulting MLEs as well as the coverage probabilities. We also provided average revision numbers needed for stabilization from each approach. We set α to zero and considered various values for β and sample size n . Our results from 10,000 Monte Carlo runs are given in Table 2.

As can be seen from Table 2, all approaches provide same biases and variances after stabilization, which is expected. The resulting coverage probabilities from all approaches are close to 0.95, and as sample size increases, both bias values and variances decrease, also as expected. For a given (α, β) value, it can be seen that the fastest stabilization is achieved by Approach 2 (i.e. equation (10)). Thus, although all three approaches yield the same results, we suggest to calculate initial $t_{(i)}$ values ($1 \leq i \leq n$) from equation (10) because the stabilization from this approach is the fastest one.

Table 2 Bias values, variances, convergence probabilities and average revision numbers using three different approaches to calculate $t_{(i)}$ values ($1 \leq i \leq n$)

			β		0.1		0.5		1.0	
n					$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\beta}$
30	Approach 1	Bias			0.0355	0.0006	0.0341	0.0067	0.0278	0.0078
		Variance			0.0373	0.0389	0.0403	0.0369	0.0451	0.0328
		Coverage prob.			0.9521	0.9537	0.9522	0.9541	0.9525	0.9488
		No of revisions			4.67		3.45		6.10	
	Approach 2	Bias			0.0355	0.0006	0.0341	0.0067	0.0279	0.0079
		Variance			0.0373	0.0388	0.0403	0.0369	0.0451	0.0328
		Coverage prob.			0.9521	0.9537	0.9522	0.9541	0.9525	0.9488
		No of revisions			2.65		2.00		1.81	
	Approach 3	Bias			0.0355	0.0006	0.0341	0.0067	0.0278	0.0079
		Variance			0.0372	0.0389	0.0402	0.0368	0.0451	0.0328
		Coverage prob.			0.9521	0.9537	0.9523	0.9540	0.9525	0.9488
		No of revisions			2.98		3.04		4.63	

Table continued...

		β	0.1		0.5		1.0	
100	Approach 1	Bias	0.0098	0.0007	0.0079	0.0002	0.0081	0.0017
		Variance	0.0104	0.0104	0.0113	0.0094	0.0126	0.0074
		Coverage prob.	0.9504	0.9518	0.9470	0.9494	0.9538	0.9495
		No of revisions	4.78		3.24		6.59	
	Approach 2	Bias	0.0098	0.0007	0.0079	0.0001	0.0081	0.0016
		Variance	0.0103	0.0104	0.0114	0.0095	0.0126	0.0073
		Coverage prob.	0.9504	0.9518	0.9470	0.9494	0.9538	0.9496
		No of revisions	2.91		2.01		1.32	
	Approach 3	Bias	0.0098	0.0007	0.0078	0.0002	0.0082	0.0017
		Variance	0.0104	0.0104	0.0113	0.0094	0.0126	0.0073
		Coverage prob.	0.9505	0.9518	0.9470	0.9494	0.9539	0.9495
		No of revisions	2.99		3.00		4.69	
250	Approach 1	Bias	0.0041	0.0002	0.0041	0.0000	0.0013	0.0007
		Variance	0.0041	0.0040	0.0044	0.0036	0.0049	0.0026
		Coverage prob.	0.9506	0.9452	0.9500	0.9479	0.9507	0.9513
		No of revisions	4.78		3.11		6.87	
	Approach 2	Bias	0.0041	0.0002	0.0040	0.0000	0.0012	0.0007
		Variance	0.0040	0.0041	0.0044	0.0036	0.0049	0.0026
		Coverage prob.	0.9506	0.9452	0.9500	0.9479	0.9507	0.9513
		No of revisions	2.90		2.00		1.12	
	Approach 3	Bias	0.0041	0.0002	0.0041	0.0001	0.0013	0.0007
		Variance	0.0041	0.0041	0.0044	0.0036	0.0049	0.0026
		Coverage prob.	0.9506	0.9452	0.9500	0.9479	0.9507	0.9513
		No of revisions	2.99		3.00		4.76	

Generalization to multivariable case

Now consider k ($k \geq 2$) covariates and assume all of them take positive values without loss of generality. The Poisson regression model with k covariates can be written as

$$E(Y_i | x_{i1}, x_{i2}, \dots, x_{ik}) = F(z_i) \quad (16)$$

where $F(z_i) = \exp(z_i)$ and

$$z_i = \beta_0 + \sum_{j=1}^k \beta_j x_{ij}, \quad (17)$$

for $1 \leq i \leq n$, $1 \leq j \leq k$. Following the same lines of,³ in order to rank the z -values we can assume that all covariates are equally effective in increasing the response Y , i.e. we initially take β_j 's all equal, and order the z -values that would correspond to the ordered x -values,

where $x_i = x_{i1} + x_{i2} + \dots + x_{ik}$ ($1 \leq i \leq n$). In other words, the ordered z -values become

$$z_{(i)} = \beta_0 + \beta_1 x_{i1}^* + \dots + \beta_k x_{ik}^*, \quad (18)$$

where the vector $\begin{bmatrix} 1 & x_{i1}^* & x_{i2}^* & \dots & x_{ik}^* \end{bmatrix}$ is the i^{th} row of the matrix

$$X^* = \begin{bmatrix} 1 & x_{11}^* & x_{12}^* & \dots & x_{1k}^* \\ 1 & x_{21}^* & x_{22}^* & \dots & x_{2k}^* \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_{n1}^* & x_{n2}^* & \dots & x_{nk}^* \end{bmatrix}, \quad (19)$$

which is constructed by arranging the rows of the X matrix

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix},$$

so as to correspond to the ordered $x_{(i)}$ value ($1 \leq i \leq n$). The MMLEs can be obtained along the same lines as in the Univariate case:

$$\hat{\Gamma} = (X^{*T} M X^*)^{-1} X^{*T} \Delta \quad (20)$$

where $\Delta = [\delta_1 \ \delta_2 \ \dots \ \delta_n]^T$, δ_i is given by (11) and M is the $n \times n$ diagonal matrix

$$M = \begin{bmatrix} b_1 & 0 & \dots & 0 \\ 0 & b_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & b_n \end{bmatrix}.$$

The asymptotic variance-covariance matrix V of the estimators can be derived from the Fisher information matrix $V = I^{-1}(\beta_0, \beta_1, \dots, \beta_k)$ as given below

$$V = \begin{bmatrix} \sum Q_i & \sum Q_i x_{1i} & \dots & \sum Q_i x_{ki} \\ \sum Q_i x_{1i} & \sum Q_i x_{1i}^2 & \dots & \sum Q_i x_{ki} x_{1i} \\ \dots & \dots & \dots & \dots \\ \sum Q_i x_{ki} & \sum Q_i x_{ki} x_{1i} & \dots & \sum Q_i x_{ki}^2 \end{bmatrix}^{-1} \quad (21)$$

where $Q_i = \exp(z_i)$. V is estimated by replacing Q_i by

$$\hat{Q}_i = \exp(\hat{z}_i), \quad \hat{z}_{(i)} = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_k x_{ik} \quad (1 \leq i \leq n).$$

Robustness

Measures of influence considered in linear regression models, such as high leverage values, are analogous in the GLM framework. Large leverage values typically mean that there are outliers in covariates. When outliers present in the data, inferences based on MLEs becomes unreliable. In fact, it has been showed that MLEs are not robust in GLMs.²¹ In Poisson regression setting, if there are outliers in the continuous covariates, the estimates can be influenced. Thus, we also studied the robustness properties of the derived MMLEs under several outlier models. We considered the Univariate model given by (1) for simplicity. We assumed that $\alpha=0$ and performed a Monte-Carlo study for different values of β , where $(n-r)$ of the observations X_1, X_2, \dots, X_n (we don't know which) come from the Standard Normal Distribution with $\sigma=1$ and the remaining r observations come from the Normal distribution with a scale $c\sigma$ where c is a positive constant. We calculated the value of r from the equation $r = [0.1n + 0.5]$ (Dixon's outlier model). The outlier models considered are:

- $(n-r)$ come from $N(0,1)$ and r come from $N(0,1)$ (No outlier situation),
- $(n-r)$ come from $N(0,1)$ and r come from $N(0,1.5)$,
- $(n-r)$ come from $N(0,1)$ and r come from $N(0,2)$,
- $(n-r)$ come from $N(0,1)$ and r come from $N(0,4)$.

Note that the model (a) above does not involve outliers and is given for the sake of comparisons. In order to be able to make direct comparisons, after generating the X values we divided them by the standard deviation of the distribution for each model. After generating the X values, we calculated $z_i = \alpha + \beta x_i$ and $\mu_i = \exp(z_i)$ for $1 \leq i \leq n$ to generate Y_i values from Poisson distribution with mean μ_i ($1 \leq i \leq n$). The values obtained from 5000 runs are given in (Table 3).

Table 3 Empirical biases and variances from dixon's outlier models

Model (a): No outlier					Model (b)				
β	n	Bias($\hat{\alpha}$)	Bias($\hat{\beta}$)	Var($\hat{\alpha}$)	Var($\hat{\beta}$)	Bias($\hat{\alpha}$)	Bias($\hat{\beta}$)	Var($\hat{\alpha}$)	Var($\hat{\beta}$)
0.1	30	0.0326	0.0075	0.0371	0.0385	0.0330	0.0021	0.0377	0.0388
	50	0.0195	0.0004	0.0214	0.0219	0.0211	0.0033	0.0214	0.0218
	100	0.0082	0.0039	0.0103	0.0102	0.0094	0.0008	0.0103	0.0104
0.2	30	0.0316	0.0020	0.0377	0.0388	0.0302	0.0034	0.0388	0.0386
	50	0.0202	0.0029	0.0217	0.0215	0.0200	0.0007	0.0219	0.0212
	100	0.0108	0.0000	0.0104	0.0102	0.0115	0.0002	0.0108	0.0103
0.4	30	0.0323	0.0033	0.0391	0.0376	0.0346	0.0012	0.0407	0.0385
	50	0.0207	0.0005	0.0224	0.0209	0.0165	0.0004	0.0230	0.0209
	100	0.0086	0.0004	0.0109	0.0097	0.0089	0.0009	0.0113	0.0099
Model (c)					Model (d)				
β	n	Bias($\hat{\alpha}$)	Bias($\hat{\beta}$)	Var($\hat{\alpha}$)	Var($\hat{\beta}$)	Bias($\hat{\alpha}$)	Bias($\hat{\beta}$)	Var($\hat{\alpha}$)	Var($\hat{\beta}$)
0.1	30	0.02827	0.0003	0.0381	0.0413	0.0385	0.0026	0.0372	0.0477
	50	0.0192	0.0013	0.0220	0.0220	0.0211	0.0004	0.0213	0.0246
	100	0.0095	0.0001	0.0105	0.0104	0.0099	0.0002	0.0108	0.0107

Table continued...

		Model (a): No outlier				Model (b)			
0.2	30	0.0362	0.0036	0.0396	0.0419	0.0350	0.0080	0.0375	0.0467
	50	0.0163	0.0014	0.0208	0.0211	0.0189	0.0057	0.0221	0.0242
	100	0.0111	0.0010	0.0106	0.0105	0.0109	0.0013	0.0099	0.0103
0.4	30	0.0313	0.0015	0.0396	0.0370	0.0314	0.0017	0.0379	0.0461
	50	0.0178	0.0030	0.0226	0.0207	0.0166	0.0001	0.0218	0.0223
	100	0.0079	0.0003	0.0110	0.0093	0.0074	0.0039	0.0111	0.0088

As can be seen from the table, the biases in the estimates are negligible for all models. The variances $\text{Var}(\hat{\beta})$ (hence the Wald statistics W) are almost the same for a given n for the models (a), (b), (c) and (d), which means that the MMLEs are robust to outliers in the covariate. Note that the MML methodology achieves robustness through the $t_{(i)}$ ($1 \leq i \leq n$) values.

Conclusion

Poisson regression serves as a useful technique to model count data. The MLEs in Poisson regression are obtained via Newton-type algorithms; however these algorithms might not converge or converge to inaccurate values. In this study we derived the explicit MMLEs for Poisson regression. We also considered the case where there are outliers in the (continuous) covariate, which generally is the case in real life applications, and searched the properties of the derived MMLEs under several data violations. Although the scope of the simulations reported here is limited, we can conclude that MML methodology provides robust estimation in Poisson regression.

Acknowledgement

We are grateful to the reviewer for the helpful comment which helped to improve this manuscript.

Conflict of interest

None.

References

1. Cameron AC, Trivedi PK. Regression Analysis of Count Data. Cambridge Univ. Press, NY, USA. 1998.
2. Casella G, Berger R. Statistical Inference. (2nd edn), Thomson Learning, Pacific Grove, CA, USA. 2002.
3. Tiku ML, Vaughan DC. Logistic and nonlogistic density functions in binary regression with nonstochastic covariates. *Biometrical Journal*. 1997;39(8):883–898.
4. Vaughan DC. The generalized secant hyperbolic distribution and its properties. *Communications in Statistics – Theory and Methods*. 2002;31(2):219–238.
5. Barnett VD. Evaluation of the maximum likelihood estimator when the likelihood equation has multiple roots. *Biometrika*. 1966;53(1):151–165.
6. Lee KR, Kapadia CH, Dwight BB. On estimating the scale parameter of Rayleigh distribution from censored samples. *Statistische Hefte*. 1980;21(1):14–20.
7. Silva JMC, Tenreiro S. Poisson: Some convergence Issues. *The Stata Journal*. 2011;11(2):207–212.
8. Silva JMC, Tenreiro S. On the Existence of the Maximum Likelihood Estimates in Poisson Regression. *Economics Letters*. 2010;107:310–312.
9. Tiku ML. Estimating the mean and standard deviation from a censored normal sample. *Biometrika*. 1967;54(1):155–165.
10. Bhattacharyya GK. The asymptotics of maximum likelihood and related estimators based on type II censored data. *Journal of the American Statistical Association*. 1985;80:398–404.
11. Tan WY. On Tiku's robust procedure—a Bayesian insight. *Journal of Statistical Planning and Inference*. 1985;11(3):329–340.
12. Tiku ML, Tan WY, Balakrishnan N. Robust Inference. Marcel Dekker, New York, USA. 1986.
13. Tiku ML, Suresh RP. A new method of estimation for location and scale parameters. *Journal of Statistical Planning and Inference*. 1992;30:281–292.
14. Vaughan DC. On the Tiku–Suresh method of estimation. *Communications in Statistics—Theory and Methods*. 1992;21(2):451–469.
15. Vaughan DC, Tiku ML. Estimation and hypothesis testing for nonnormal bivariate distribution with applications. *Mathematical and Computer Modelling*. 2000;32:53–67.
16. Oral E, Gunay S. Stochastic Covariates in Binary Regression. *Hacettepe Journal of Mathematics and Statistics*. 2004;33:97–109.
17. Oral E. Binary Regression with Stochastic Covariates. *Communications in Statistics Theory and Methods*. 2006;35:1429–1447.
18. Oral E. Parameter Estimation in Generalized Linear Models through Modified Maximum Likelihood. International Statistical Institute. Proceedings of the 58th World Statistical Congress. 2011.
19. Agresti A. Categorical Data Analysis. John Wiley and Sons, New York, USA. 1996.
20. Kwan R, Lee, Kapadia C H, Dwight B. et al. On Estimating the Scale Parameter of the Rayleigh distribution from Doubly Censored Samples. *Statistische Hefte*. 1980;21(1):14–29.
21. Cantoni E, Ronchetti E. Robust Inference for Generalized Linear Models. *Journal of the American Statistical Association*. 2001;96(455):1022–1030.