

Robust Estimators for Poisson Regression

Idriss Abdelmajid Idriss, Weihu Cheng

College of Applied Science, Department of Statistics, Beijing University of Technology, Beijing, China

Email: Idrissabdo59@gmail.com, Chengweihu@bjut.edu.cn

How to cite this paper: Idriss, I.A. and Cheng, W.H. (2023) Robust Estimators for Poisson Regression. *Open Journal of Statistics*, 13, 112-118.

<https://doi.org/10.4236/ojs.2023.131007>

Received: December 27, 2022

Accepted: February 25, 2023

Published: February 28, 2023

Copyright © 2023 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The present paper proposes a new robust estimator for Poisson regression models. We used the weighted maximum likelihood estimators which are regarded as Mallows-type estimators. We perform a Monte Carlo simulation study to assess the performance of a suggested estimator compared to the maximum likelihood estimator and some robust methods. The result shows that, in general, all robust methods in this paper perform better than the classical maximum likelihood estimators when the model contains outliers. The proposed estimators showed the best performance compared to other robust estimators.

Keywords

Poisson Regression Model, Maximum Likelihood Estimator, Robust Estimation, Contaminated Model, Weighted Maximum Likelihood Estimator

1. Introduction

Poisson regression model is widely used for modeling response variables that are counted. It is discussed by [1]. Practically, a common method used to estimate parameters is the maximum likelihood estimator (MLE). Unfortunately, this technique is high sensitivity to outliers in data, see ([2] [3]). To overcome this issue, many robust are an alternative to Maximum Likelihood Estimates. One of the first robust methods used to estimate the parameters in Poisson regression models is the Conditionally Unbiased Bounded Influence introduced by [4]. [5] discussed the M-estimator for Poisson regression model, these estimates belong to Mallows-type. [6] developed robust M-estimates for generalized linear models (GLM), these estimates are asymptotically normal and consistent. [7] proposed a fast and stable technique based on breakdown point of the trimmed maximum likelihood for generalized linear models. [8] introduced the class of M-estimators

based on quasi likelihood estimators proposed by [9]. [10] developed a robust estimator for Poisson regression model based on Mallows quasi-likelihood estimator. [11] discussed the behaviour of maximum likelihood estimator in the present of outliers. [12] discussed a robust resistant estimator based on the misclassification model. [13] generalized Optimally Bounded Score Function discussed by [14] for linear models to the generalized linear model. More recently, [15] introduced a robust method for logistic regression models. [16] discussed the robust estimators for Poisson regression model with outliers.

In this paper, we introduced a robust method for Poisson regression by using the weight functions proposed by [17] these weight functions are based on Mallows's type estimator, moreover, to evaluate the performance of the new methods with (MLE), Mallows, and (CUBIF) using the Monte Carlo simulation study. In Section 2, we discuss the Poisson regression model and the maximum likelihood estimators. In Section 3, we provide robust estimators for Poisson regression. In Section 4, we show the results of Monte Carlo simulation study. In Section 5, we offer the conclusions.

2. Poisson Regression Model and ML Estimator

Poisson regression is proper method to model a count data. Probability mass function is

$$P(Y = y) = \frac{\exp^{-\mu} \mu^y}{y!}; y = 0, 1, 2, \dots \quad (1)$$

where: $E(Y) = \mu$ and $Var(Y) = \mu$, that is mean the Poisson regression has equal mean and variance. Based on a sample (y_1, y_2, \dots, y_n) , the Poisson regression model in terms of the mean of response can be written as follow: $E(y_i) = \mu$

$$y_i = E(y_i) + \varepsilon_i, i = 1, 2, \dots, n, \quad (2)$$

where ε_i are disturbance terms. The relationship between the mean of the dependent variable and explanatory variable can be describe by use the log-link function:

$$\mu_i = e^{x_i^T \beta}, \quad (3)$$

where $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})$ is the explanatory variables and $\beta = (\beta_1, \beta_2, \dots, \beta_p)$ is the parameters of regression. The popular method used to estimate parameters in Poisson regression models is the maximum likelihood estimation, the likelihood function of the response variables (y_1, y_2, \dots, y_n) is:

$$l(Y, \mu) = \prod_{i=1}^n p_i(y_i) = \prod_{i=1}^n \frac{\exp^{-\mu_i} \mu_i^{y_i}}{y_i!}, \quad (4)$$

$$\log l(Y, \mu) = \sum_{i=1}^n y_i \log(\mu_i) - \sum_{i=1}^n \mu_i - \sum_{i=1}^n \lg(y_i!)$$

$\log l(y_1, y_2, \dots, y_n) | \beta, x_1, x_2, \dots, x_n = \sum_{i=1}^n \log p(Y_i = y | \beta, x_i)$, with according

to (1), $p(Y_i = y_i | \beta, x_i) = \frac{\exp(-\mu_i) \mu_i^{y_i}}{y_i!}$ and $\mu_i = \exp(\beta^T x_i)$. We can write the total likelihood as follows:

$$\log l(\beta) = \sum_{i=1}^n \left[-\exp(\beta^T x_i) - \log(y_i!) \right]$$

We can define the MLE as: $\hat{\beta}_{ML} = \arg \max l(\beta)$. To get the estimates of maximum likelihood for this model, we can maximize the likelihood function by differentiating it respect to β . While, maximizing the likelihood function has no closed form solution so may use the (Fisher Scoring) or the iteratively weighted least squares algorithm (IWLS) to get the maximum likelihood estimates see ([1] [18]).

In this paper, we focus on maximum weighted likelihood estimators for Poisson regression model, the maximum weighted likelihood estimator is:

$$\hat{\beta}_{ML} = (X^T \hat{W} X)^{-1} X^T \hat{W} \hat{Z}$$

There $\hat{Z} = (\hat{Z}_1, \hat{Z}_2, \dots, \hat{Z}_n)^T$, with

$$\hat{Z}_i = \log(\hat{\mu}_i) + \frac{y_i - \hat{\mu}_i}{\hat{\mu}_i}, \quad \hat{W}_i = \text{diag}(\hat{\mu}_i), \quad \text{and} \quad X = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix}$$

3. Robust Poisson Regression

In robust Poisson regression, Mallows-type estimator introduced by [4] can be applied to fit the data a count variables, this method minimizes the weighted log-likelihood function. [5] studied Mallows-type estimator deeply and introduced a robust method for generalized linear models. We can measure the leverage of observation x by used the following:

$$h_n(x) = \left((x - \hat{\mu}_n)^T \hat{\Sigma}_n^{-1} (x - \hat{\mu}_n) \right)^{1/2}, \quad (5)$$

where $\hat{\mu}_n$ is the robust location estimator and $\hat{\Sigma}_n$ is the robcation estimator $\hat{\Sigma}$ and $\hat{\mu}$, can be calculated by using minimum covariance determinate (MCD) method. We can get the Mallows-type estimator fust variance-covariance matrix of the predictor variables (x_1, x_2, \dots, x_n) . The robust scale and loor Poisson regression by solution the equations:

$$\sum_{i=1}^n w_i \left[y_i \log(\mu_i) - \mu_i - \log(y_i!) \right], \quad (6)$$

where $w_i = w(h_n(x_i))$, w is a non increasing function such that $w(u)$ is bounded. [5] introduced choosing w depends on a constant $c > 0$.

$$W(u) = \left(1 - \frac{u^2}{c^2} \right)^3 I(|u| \leq c),$$

this estimate knows as Mallows-type estimator or weighted maximum likelihood estimator (WMLE).

In this paper, we introduced a robust methods for Poisson regression model they are based on maximum weighted likelihood estimators. The weight of this method depends on the function introduced by [17]. We first calculate the initial scatter and location estimators of predictor $\hat{\Sigma}^{(0)}$ and $\hat{\mu}^{(0)}$ respectively. then, compute the squared Mahalanobis distances of predictor which can be defined as:

$$m^2 = (x_i - \hat{\mu}^{(0)})^T (\hat{\Sigma}^{(0)})^{-1} (x_i - \hat{\mu}^{(0)}).$$

The weight function we introducing can be defined as follows: first weight: $w_1 = (0.8 * m^2 + 0.2)$, where m^2 indicate to squares Mahalanobis distances, then:

$$w_1 = \left(0.8 * (x_i - \hat{\mu}^{(0)})^T (\hat{\Sigma}^{(0)})^{-1} (x_i - \hat{\mu}^{(0)}) + 0.2 \right),$$

second weight: $w_2 = (0.8 * (m^2)^2 + 0.2)$, then, we can write in the form of:

$$w_2 = \left(0.8 * \left((x_i - \hat{\mu}^{(0)})^T (\hat{\Sigma}^{(0)})^{-1} (x_i - \hat{\mu}^{(0)}) \right)^2 + 0.2 \right).$$

Then, the maximum weighted likelihood estimators for Poisson regression model can be gained by a solution the following form:

$$\sum_{i=1}^n w_i \left[y_i \log(\mu_i) - \mu_i - \log(y_i!) \right]. \quad (7)$$

For compute the maximum weighted likelihood estimators we used algorithm of Mallows-type introduced by [5].

4. Evaluation of the Robust Methods

In order to test the performance of the above estimates, we conducted Monte Carlo simulation study for comparing the a new methods with the Maximum likelihood estimator (MLE), Mallows type estimator for [4] and [5].

4.1. Monte Carlo Simulation Study

In this subsection, we examine the perform of the new robust methods ($WMLEw_1$, $WMLEw_2$) and compare with the maximum likelihood estimate (MLE), Mallows-type estimator (Mallows) of [5] and the conditionally unbiased bounded influence (CUBI) of [4]. The simulation study includes three models. First model is clean model, second model is 5% of data contaminated and third model is 10% of data are contaminated. In the three models, we generated the explanatory variables x_i from standard normal distribution $N_p(0,1)$, and the victor of parameter is $\beta(1,2,2)$, with four sample size $n = (100, 200, 300, 400)$, these values were chose to represent moderate and large samples.

The response variables y_i are generated from poisson distribution with $p(\mu_i)$ with $\mu_i = h(\eta_i)$. The outliers are distributed according to Poisson distribution with mean $3IQR(e^{x\beta})$: where IQR is the interquartile range. To examine the

perform of these estimators, we compute the Bias and mean squared error (MSE) for the three models. For all scenarios, we run 1000 repetitions. However, a good estimator is the one has small Bias and MSE. Therefore, we compute the bias and MSE for each parameter as follows:

$$\text{Bias} = \left| \frac{1}{1000} \sum_{i=1}^{1000} \beta_i - \beta \right|,$$

and

$$\text{MSE} = \frac{1}{1000} \sum_{i=1}^n |\hat{\beta}_i - \beta|^2.$$

4.2. Results from the Monte Carlo Simulation Study

It is seen in **Table 1** for clean model, the values of the Bias and MSE for maximum likelihood estimator (MLE), (Mallows) and (CUBIF) are smaller those of new weighted estimators ($WMLE_1$), ($WMLE_2$). We can conclude that the $WMLE_1$ and $WMLE_2$ estimators perform less compared to others in clean models. But when the 5% of data are contaminated in second scenario (**Table 2**) and 10% of data are contaminated in third scenario (**Table 3**), the weighted maximum likelihood estimators ($WMLE_1$ and $WMLE_2$) has lower bias and (MSE) compered with others methods, that is mean our new methods perform better compered with other estimators.

Table 1. Bias and MSE of estimators for clean model.

Methods	$n = 100$		$n = 200$		$n = 300$		$n = 400$	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
MLE	0.1572	0.0104	0.1392	0.0231	0.1169	0.0233	0.0958	0.0201
MALLOWS	0.1574	0.0109	0.1389	0.0240	0.0970	0.0224	0.0983	0.0192
CUBIF	0.2027	0.0501	0.1225	0.0207	0.0917	0.0210	0.0915	0.0197
$WMLE_1$	0.2739	0.8159	0.3014	0.2572	0.2014	0.5790	0.2148	0.0687
$WMLE_2$	0.2545	0.7132	0.4142	0.3491	0.1595	0.6781	0.1810	0.0799

Table 2. Bias and MSE of estimators when 5% of data are contaminated.

Methods	$n = 100$		$n = 200$		$n = 300$		$n = 400$	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
MLE	0.7160	1.3112	0.6871	1.3401	0.7071	1.12289	0.7082	1.1534
MALLOWS	0.7481	1.3032	0.6158	1.3105	0.7506	1.4283	0.6857	1.1342
CUBIF	0.6378	1.3346	0.5229	1.2176	0.6918	1.2111	0.6014	1.1796
$WMLE_1$	0.2571	0.1470	0.1622	0.11645	0.2116	0.2269	0.3285	0.2027
$WMLE_2$	0.3294	0.1437	0.2013	0.1012	0.2135	0.3453	0.2209	.1972

Table 3. Bias and MSE of estimators when 10% of data are contaminated.

Methods	$n = 100$		$n = 200$		$n = 300$		$n = 400$	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
MLE	0.7816	1.2201	0.6687	1.3046	0.6870	1.2101	0.6957	1.1815
MALLOWS	0.7925	1.1978	0.5088	1.2942	0.7305	1.3028	0.7034	1.1624
CUBIF	0.6836	1.1844	0.5829	1.2217	0.7028	1.1826	0.6314	1.2007
$WMLE_1$	0.3125	0.4723	0.0862	0.9165	0.9614	0.1926	0.3018	0.1852
$WMLE_2$	0.3092	0.3725	0.1322	0.2412	0.2113	0.2743	0.2120	0.1723

5. Conclusion

In this paper, we suggested new robust estimators for Poisson regression: the weighted maximum likelihood estimators ($WMLE_1$ and $WMLE_2$). To examine the performance of suggested estimators, we conducted a Monte Carlo simulation study to compare the suggested estimators with the classical maximum likelihood estimator (ML), Mallows and CUBIF. The result of simulation study shows that in **Table 1** (clean model), the maximum likelihood estimator, CUBIF and Mallows perform close to each other, while the proposed estimators have lower performance compared to other estimators. In contaminated models (**Table 2** and **Table 3**), the new weighted maximum likelihood estimators ($WMLE_1$ and $WMLE_2$) performs better compared with other estimators. The extent of new estimators proposed in this paper to other generalized linear models would be an interesting subject to follow.

Acknowledgements

We thank the editor and the referee for their comments.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] McCullagh, P. and Nelder, J.A. (1989) Generalized Linear Models. Chapman Hall, London. <https://doi.org/10.1007/978-1-4899-3242-6>
- [2] Agresti, A. (2001) Categorical Data Analysis. Wiley, New York. <https://doi.org/10.1002/0471249688>
- [3] Cameron, A. and Trivedi, P. (2013) Regression Analysis of Count Data. Cambridge University Press, Cambridge. <https://doi.org/10.1017/CBO9781139013567>
- [4] Kunsch, H.R. and Stefonski, L.A. and Carroll, R.J. (1989) Conditionally Unbiased Bounded-Influence Estimation in General Regression Models, with Applications to Generalized Linear Models. *Journal of the American Statistical Association*, **84**, 460-466. <https://doi.org/10.1080/01621459.1989.10478791>
- [5] Carroll, R.J. and Pederson, S. (1993) On Robustness in the Logistic Regression

- Model. *Journal of the Royal Statistical Society. Series B (Methodological)*, **55**, 693-706. <https://doi.org/10.1111/j.2517-6161.1993.tb01934.x>
- [6] Bianco, A.M. and Yohai, V.J. (1996) Robust Estimation in the Logistic Regression Model. In: Reider, H., Ed., *Robust Statistics, Data Analysis, and Computer Intensive Methods*, Springer, Berlin, 17-34. https://doi.org/10.1007/978-1-4612-2380-1_2
- [7] Croux, C. and Haesbroeck, G. (2003) Implementing the Bianco and Yohai Estimator for Logistic Regression. *Computational Statistics and Data Analysis Journal*, **44**, 273-295. [https://doi.org/10.1016/S0167-9473\(03\)00042-2](https://doi.org/10.1016/S0167-9473(03)00042-2)
- [8] Cantoni, E. and Ronchetti, E. (2001) Robust Inference for Generalized Linear Models. *Journal of the American Statistical Association*, **544**, 1022-1030. <https://doi.org/10.1198/016214501753209004>
- [9] Wedderburn, R.W.M. (1974) Quasi-Likelihood Functions, Generalized Linear Models, and the Gauss Newton Method. *Biometrika*, **61**, 439-447. <https://doi.org/10.1093/biomet/61.3.439>
- [10] Hosseinian, S. and Morgenthaler, S. (2011) Weighted Maximum Likelihood Estimates in Poisson Regression. *International Conference on Robust Statistics*, Antalya, 2008.
- [11] Croux, C., Flandre, C. and Haesbroeck, G. (2002) The Breakdown Behavior of the Maximum Likelihood Estimator in the Logistic Regression Model. *Statistics and Probability Letters*, **60**, 377-386. [https://doi.org/10.1016/S0167-7152\(02\)00292-4](https://doi.org/10.1016/S0167-7152(02)00292-4)
- [12] Copas, J.B. (1988) Binary Regression Models for Contaminated Data. *Journal of the Royal Statistical Society B*, **50**, 225-265. <https://doi.org/10.1111/j.2517-6161.1988.tb01723.x>
- [13] Stefanski, L.A., Carroll, R.J. and Ruppert, D. (1986) Optimally Bounded Score Functions for Generalized Linear Models with Applications to Logistic Regression. *Biometrika*, **2**, 413-424. <https://doi.org/10.1093/biomet/73.2.413>
- [14] Krasker, W.S. and Welsch, R.E. (1982) Efficient Bounded-Influence Regression Estimation. *Journal of the American Statistical Association*, **77**, 595-604. <https://doi.org/10.1080/01621459.1982.10477855>
- [15] Ahmed, I.A.I. and Cheng, W.H. (2020) The Performance of Robust Methods in Logistic Regression Model. *Open Journal of Statistics*, **10**, 127-138. <https://doi.org/10.4236/ojs.2020.101010>
- [16] Abonazel, M. and Saber, O. (2020) A Comparative Study of Robust Estimators for Poisson Regression Model with Outliers. *Journal of Statistics Applications and Probability*, **2**, 279-286. <https://doi.org/10.18576/jsap/090208>
- [17] Simeckova, M. (2005) Maximum Weighted Likelihood Estimator in Logistic Regression. Charles University, Faculty of Mathematics and Physics, Prague, 144-148.
- [18] Dobson, A. and Barnett, A. (2008) An Introduction to Generalized Linear Models. Chapman, Hall. <https://doi.org/10.1201/9780367807849>