International Journal of Mathematics and
Computer Science, **19**(2024), no. 4, 933–939

Fisher Information Matrix for Generalized Poisson Regression: Evaluation of the Log-Likelihood Function

Riski Nur Istiqomah Dinnullah[1,2], **Sobri Abusini**[1], **Rahma Fitriani**[1], **Marjono**[1], **Trija Fayeldi**[2], **Rauzan Sumara**[3]

[1]University of Brawijaya
Jl. Veteran, Ketawanggede, Malang City, Indonesia

[2]Universitas PGRI Kanjuruhan Malang
Jl. S. Supriadi 48, Malang City, Indonesia

[3]Warsaw University of Technology
plac Politechniki 1, 00-661 Warszawa, Polandia

email: ky2_zahra@unikama.ac.id

(Received January 8, 2024, Accepted February 19, 2024,
Published June 1, 2024)

Abstract

Fisher information is an essential element in statistical modeling and is required for a matrix-based parameter estimator to find the optimal solution. The information matrix is calculated by subtracting the expectation value matrix of the function to be maximized by a given amount. Positive semidefiniteness is observed in this matrix with regard to each parameter value. The Fisher information matrix (FIM) shows how parameters in a probabilistic model are related to each other. It is an inherent consequence of the procedure of maximum likelihood estimation (MLE). In this paper, we perform an analytical evaluation of the FIM for Generalized Poisson Regression (GPR). In the previous stage, we analyzed the expectation of the second derivative, where the evaluation function is the log-likelihood function for the GPR model.

Key words and phrases: Fisher Information Matrix, Generalized Poisson Regression, Log-likelihood function.
AMS (MOS) Subject Classifications: 60E05, 60E10.
The corresponding author is Riski Nur Istiqomah Dinnullah.
ISSN 1814-0432, 2024, http://ijmcs.future-in-tech.net

1 Introduction

Critical to statistical testing, the FIM is utilized to identify optimal experimental conditions, compute the Wald test statistic, and ascertain the classical asymptotic distribution [1]. It also facilitates the evaluation of parameter estimate accuracy at infinity. In addition, the researchers have employed the FIM to examine specific Riemannian metrics on complex manifolds [2], to deduce post-model selection [3], and to ascertain the asymptotic distribution for the variance components in the mixed model [4]. The precise calculation, however, is frequently not trivial. Latent variable models, which include unobserved variables and are also known as incomplete data models, pose a significant challenge for precise calculation. Despite the fact that these models are becoming more prevalent in numerous application domains, including genomics [5], ecology [6], and ecophysiology [7], this facilitates a more comprehensive analysis of diverse sources of variability and, if necessary, a more precise identification of the established mechanisms that underlie the data.

The Fisher information matrix shows how parameters in a probabilistic model are related to each other. It is a natural result of the maximum likelihood estimation (MLE) process [8]. Thus the purpose of this study is to perform an analytical evaluation of the FIM for Generalized Poisson Regression based on log-likelihood function. In the previous stage, we analyzed the expectation of the second derivative, where the evaluation function is the log-likelihood function for the GPR model.

2 Generalized Poisson Regression

In accordance with the following probability function, a set of data is said to follow the Generalized Poisson Distribution:

$$p\left(y_i;\psi_i,\varsigma\right)=\left(\frac{\psi_i}{\varsigma\psi_i+1}\right)^{y_i}\frac{\left(\varsigma y_i+1\right)^{y_i-1}}{y_i!}e^{-\frac{\psi_i\left(\varsigma y_i+1\right)}{\varsigma\psi_i+1}},i=1,2,3,...,n, \tag{2.1}$$

where $\psi_i=e^{x\beta}=e^{\beta_0+\sum\limits_{j=1}\beta_j x_{ji}}$ [9].

3 Maximum Likelihood Estimation (MLE)

Parameter estimation with MLE is a dependable and potent statistical technique [10, 11]. The log-likelihood function is an essential component of the

MLE procedure. GPR's log-likelihood function is provided for the purposes of this article.

$$\iota(\boldsymbol{\beta},\varsigma)=\sum_{i=1}^{n}\left(\left(\beta_0+\sum_{j=1}^{p}\beta_j x_{ji}\right)y_i-\ln\left(\varsigma\exp\left(\beta_0+\sum_{j=1}^{p}\beta_j x_{ji}\right)+1\right)y_i+(y_i-1)\right.$$
$$\left.\ln\left(\varsigma y_i+1\right)-\frac{(\varsigma y_i+1)}{\varsigma\exp\left(\beta_0+\sum\limits_{j=1}^{p}\beta_j x_{ji}\right)+1}\exp\left(\beta_0+\sum_{j=1}^{p}\beta_j x_{ji}\right)-\ln(y_i!)\right). \tag{3.2}$$

First derivative calculation is necessary for the log likelihood function to be maximized. Next, several theorems are presented as analytical justifications by analysis of the expected value of second derivative to build FIM. For a special case, when the expected value of the second partial derivative of the parameter must be determined, the $E\left[\frac{y_i^2[y_i-1]}{(1+\varsigma y_i)^2}\right]$ value is ascertained for special events, using the definition of moments and recurrence relations.

Theorem 3.1. *Suppose that the second derivative of β_0 of equation 3.2 is known. Then*

$$E\left[\frac{\partial^2\iota\left(\boldsymbol{\beta},\varsigma\right)}{\partial{\beta_0}^2}\right]=\sum_{i=1}^{n}\frac{-\psi_i}{\left(\varsigma\psi_i+1\right)^3}. \tag{3.3}$$

Proof.

$$E\left[\frac{\partial^2\iota\left(\boldsymbol{\beta},\varsigma\right)}{\partial{\beta_0}^2}\right]=\sum_{i=1}^{n}\left(E\left(\frac{\varsigma\psi_i^2}{\left(\varsigma\psi_i+1\right)^3}\right)-E\left(\frac{2\varsigma\psi_i y_i}{\left(\varsigma\psi_i+1\right)^3}\right)+E\left(\frac{-\psi_i}{\left(\varsigma\psi_i+1\right)^3}\right)\right)$$
$$=\sum_{i=1}^{n}\frac{-\psi_i+\varsigma\psi_i^2-2\varsigma\psi_i^2}{\left(1+\varsigma\psi_i\right)^3}=\sum_{i=1}^{n}\frac{-\psi_i\left(1+\varsigma\psi_i\right)}{\left(1+\varsigma\psi_i\right)^3}=\sum_{i=1}^{n}\frac{-\psi_i}{\left(\varsigma\psi_i+1\right)^2}.$$

□

Theorem 3.2. *Suppose that the second derivative of β_k of equation 3.2 is known. Then*

$$E\left[\frac{\partial^2\iota\left(\boldsymbol{\beta},\varsigma\right)}{\partial{\beta_k}^2}\right]=\sum_{i=1}^{n}\frac{-\left(x_{ki}\right)^2\psi_i}{\left(\varsigma\psi_i+1\right)^3},\quad \textit{for } k=1,2,3,\ldots,p. \tag{3.4}$$

Proof. The proof is similar to that of theorem 3.1. □

Theorem 3.3. *Suppose that the second derivative of $\beta_m\beta_0$ or $\beta_k\beta_0$ of equation 3.2 is known. Then*

$$E\left[\frac{\partial^2\iota\left(\boldsymbol{\beta},\varsigma\right)}{\partial\beta_0\partial\beta_m}\right]=E\left[\frac{\partial^2\iota\left(\boldsymbol{\beta},\varsigma\right)}{\partial\beta_k\partial\beta_0}\right]=\sum_{i=1}^{n}\frac{-x_{mi}\psi_i}{\left(\varsigma\psi_i+1\right)^2},\quad \textit{for } m=k=1,2,3,\ldots,p. \tag{3.5}$$

Proof. The proof is similar to that of theorem 3.1. □

Theorem 3.4. *Suppose that the second derivative of $\beta_m\beta_k$ of equation 3.2 is known. Then*

$$E\left[\frac{\partial^2 \iota\left(\boldsymbol{\beta},\varsigma\right)}{\partial\beta_k\partial\beta_m}\right]=\sum_{i=1}^{n}\frac{-x_{mi}x_{ki}\psi_i}{\left(\varsigma\psi_i+1\right)^2}, \tag{3.6}$$

for $m=1,2,3,...,p$ and $k=1,2,3,...,p$, with $m\neq k$.

Proof. The proof is similar to that of theorem 3.1. □

Theorem 3.5. *Suppose that the second derivative of $\varsigma\beta_0$ of equation 3.2 is known. Then*

$$E\left[\frac{\partial^2 \iota\left(\boldsymbol{\beta},\varsigma\right)}{\partial\beta_0\partial\varsigma}\right]=E\left[\frac{\partial^2 \iota\left(\boldsymbol{\beta},\varsigma\right)}{\partial\varsigma\partial\beta_0}\right]=0. \tag{3.7}$$

Proof. The proof is similar to that of theorem 3.1. □

Theorem 3.6. *Suppose that the second derivative of $\beta_m\varsigma$ or $\beta_k\varsigma$ of equation 3.2 is known. Then*

$$E\left[\frac{\partial^2 \iota\left(\boldsymbol{\beta},\varsigma\right)}{\partial\varsigma\partial\beta_m}\right]=E\left[\frac{\partial^2 \iota\left(\boldsymbol{\beta},\varsigma\right)}{\partial\beta_k\partial\varsigma}\right]=0, m=k=1,2,3,...,p. \tag{3.8}$$

Proof. The proof is similar to that of theorem 3.1. □

Then, to get the expectation of the second derivative of ς, the $E\left[\frac{y_i^2(y_i-1)}{(\varsigma y_i+1)^2}\right]$ value is determined using the definition of moment and recurrence relations.

Theorem 3.7. *Let $\mu_k=E\left[\left(\frac{y_i^2(y_i-1)}{(\varsigma y_i+1)^2}\right)^k\right]$. Then the recurrence relation is*

$$\left(2k-1\right)\mu_{k+1}=-\frac{\psi_i^2}{\left(\varsigma\psi_i+1\right)}\mu_k-\frac{\partial\mu_k}{\partial\varsigma}-\left(2\psi_i^2+\varsigma\psi_i^3\right)\frac{\partial\mu_k}{\partial\psi_i}. \tag{3.9}$$

Proof. The k-th moment of $\left(\frac{y_i^2(y_i-1)}{(\varsigma y_i+1)^2}\right)^k$ can be written as follows:

$$\mu_k=E\left[\left(\frac{y_i^2\left(y_i-1\right)^2}{\left(\varsigma y_i+1\right)}\right)^k\right]=\sum_{y}\left(\frac{y_i^2\left(y_i-1\right)}{\left(\varsigma y_i+1\right)^2}\right)^k\left(\frac{\psi_i}{\varsigma\psi_i+1}\right)^{y_i}\frac{\left(\varsigma y_i+1\right)^{y_i-1}}{y_i!}e^{-\frac{\psi_i\left(\varsigma y_i+1\right)}{\varsigma\psi_i+1}} \tag{3.10}$$

Then the next step is to look for the derivative of μ_k with respect to ψ_i, and μ_k with respect to ς. By the substitution process, we obtain the equation

$$\frac{\partial \mu_k}{\partial \varsigma} = (-2k+1)\,\psi_i^{k+1} + \frac{-2\psi_i - \varsigma\psi_i^2}{(\varsigma\psi_i+1)^2}\left(\psi_i\,(\varsigma\psi_i+1)^2\,\frac{\partial \mu_k}{\partial \psi_i} + \psi_i\mu_k\right) + \frac{\psi_i^2}{(\varsigma\psi_i+1)^2}\mu_k. \tag{3.11}$$

Thus we have proven that $(2k-1)\,\mu_{k+1} = -\frac{\psi_i^2}{(\varsigma\psi_i+1)}\mu_k - \frac{\partial\mu_k}{\partial\varsigma} - (2\psi_i^2 + \varsigma\psi_i^3)\,\frac{\partial\mu_k}{\partial\psi_i}$. Next, we can find $E\left[\frac{y_i^2(y_i-1)}{(\varsigma y_i+1)^2}\right]$ by substituting $k=0$ in equation 3.9 and the result is

$$E\left[\frac{y_i^2\,(y_i-1)}{(\varsigma y_i+1)^2}\right] = \mu_1 = \frac{\psi_i^2}{\varsigma\psi_i+1}. \tag{3.12}$$

□

Theorem 3.8. *Suppose the second derivative of ς of equation 3.2 is known. Then*

$$E\left[\frac{\partial^2\iota\,(\boldsymbol{\beta},\varsigma)}{\partial\varsigma^2}\right] = \sum_{i=1}^{n}\left(\frac{\psi_i^3}{(\varsigma\psi_i+1)^2} - \frac{\psi_i^2}{(\varsigma\psi_i+1)}\right). \tag{3.13}$$

Proof. The proof is similar to that of theorem 3.1 and substituting equation 3.12 into the solution.

Using the expected value of the second derivative, the FIM may be expressed as follows:

$$I\,(\breve{\gamma}) = \begin{bmatrix} \sum_{i=1}^{n}\frac{\mu_i}{(\varsigma\psi_i+1)^2} & \sum_{i=1}^{n}\frac{x_{1i}\psi_i}{(\varsigma\psi_i+1)^2} & \sum_{i=1}^{n}\frac{x_{2i}\psi_i}{(\varsigma\psi_i+1)^2} & \cdots & \sum_{i=1}^{n}\frac{x_{pi}\psi_i}{(\varsigma\psi_i+1)^2} & 0 \\ \sum_{i=1}^{n}\frac{x_{1i}\psi_i}{(\varsigma\psi_i+1)^2} & \sum_{i=1}^{n}\frac{(x_{1i})^2\psi_i}{(\varsigma\psi_i+1)^2} & \sum_{i=1}^{n}\frac{x_{1i}x_{2i}\psi_i}{(\varsigma\psi_i+1)^2} & \cdots & \sum_{i=1}^{n}\frac{x_{1i}x_{pi}\psi_i}{(\varsigma\psi_i+1)^2} & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \sum_{i=1}^{n}\frac{x_{pi}\psi_i}{(\varsigma\psi_i+1)^2} & \sum_{i=1}^{n}\frac{(x_{pi})^2\psi_i}{(\varsigma\psi_i+1)^2} & \sum_{i=1}^{n}\frac{x_{pi}x_{2i}\psi_i}{(\varsigma\psi_i+1)^2} & \cdots & \sum_{i=1}^{n}\frac{(x_{pi})^2\psi_i}{(\varsigma\psi_i+1)^2} & 0 \\ 0 & 0 & 0 & \cdots & 0 & \sum_{i=1}^{n}\left(\frac{\psi_i^2}{(\varsigma\psi_i+1)} - \frac{\psi_i^3}{(\varsigma\psi_i+1)^2}\right) \end{bmatrix}. \tag{3.14}$$

□

The FIM is key in statistical modeling and is especially effective for estimation problems when the MLE is nonlinear in form which makes things very difficult to solve analytically. Therefore, the Fisher Information Matrix is very useful when used in the numerical optimization of the Fisher-Scoring Algorithm. Therefore, further research can utilize the Fisher Information Matrix in algorithms and case studies to gain a clearer understanding of its performance in applied scenarios.

References

[1] W. Liu, Y. Wang, W. Xie, Fisher information matrix, Rao test, and Wald test for complex-valued signals and their applications, Signal Process, **94,** (2014), 1–5, https://doi.org/10.1016/j.sigpro.2013.06.032.

[2] A. Le Brigant, S. C. Preston, S. Puechmorel, Fisher-Rao geometry of Dirichlet distributions, Differential Geometry and its Applications, **74,** (2021), 101702, https://doi.org/10.48550/arXiv.2005.05608.

[3] A. Charkhi, G. Claeskens, Asymptotic post-selection inference for the Akaike information criterion, Biometrika, **105,** no. 3, (2018), 645–664, https://doi.org/10.1093/biomet/asy018.

[4] C. Baey, P. H. Courn'ede, E. Kuhn, Asymptotic distribution of likelihood ratio test statistics for variance components in nonlinear mixed effects models, Computational Statistics and Data Analysis, **135,** (2019), 107–122, https://doi.org/10.1016/j.csda.2019.01.014.

[5] F. Picard, S. Robin, E. Lebarbier, J. J. Daudin, A Segmentation/Clustering Model for the Analysis of Array CGH Data, Biometrics, **63,** (2007), 758–766, https://doi.org/10.1111/j.1541-0420.2006.00729.x.

[6] P. Gloaguen, S. Mahevas, E. Rivot, M. Woillez, J. Guitton, Y. Vermard, M. P. Etienne, An autoregressive model to describe fishing vessel movement and activity, Environmetrics, **26,** no. 1, (2014), 17–28, http://dx.doi.org/10.1002/env.2319.

[7] F. Technow, C. D. Messina, L. R. Totir, M. Cooper, Integrating crop growth models with whole genome prediction through approximate Bayesian computation, PloS one, **10,** no. 6, (2015), e0130855, https://doi.org/10.1371/journal.pone.0130855.

[8] M. Delattre, E. Kuhn, Estimating Fisher Information Matrix in Latent Variable Models based on the Score Function, (2023), arXiv:1909.06094, https://doi.org/10.48550/arXiv.1909.06094.

[9] F. Famoye, Restricted Generalized Poisson Regression model, Comm. Stat. Theory Methods, **22,** no. 5, (1993), 1335–1354, https://doi.org/10.1080/03610929308831089.

[10] A. Prahutama, D. Ispriyanti, B. Warsito, Modelling Generalized Poisson Regression in the number of Dengue Hemorrhagic Fever (DHF) in East Nusa Tenggara, E3S Web of Conferences, **202,** (2020), 12017, https://doi.org/10.1051/e3sconf/202020212017.

[11] R. N. I. Dinnullah, S. Abusini, R. Fitriani, Marjono, Application of Generalized Poisson Regression for the development of modelling motorcycle accidents based on rider characteristics, Journal of Physics: Conference Series, 2634, no. 1, (2023), 012007, https://iopscience.iop.org/article/10.1088/1742-6596/2634/1/012007.