

Optimal designs for small Poisson regression experiments using second-order asymptotic

S. Mehr Mansour^a, M. Niaparast^{1,a}

^aRazi University, Department of Statistics, Iran

Abstract

This paper considers the issue of obtaining the optimal design in Poisson regression model when the sample size is small. Poisson regression model is widely used for the analysis of count data. Asymptotic theory provides the basis for making inference on the parameters in this model. However, for small size experiments, asymptotic approximations, such as unbiasedness, may not be valid. Therefore, first, we employ the second order expansion of the bias of the maximum likelihood estimator (MLE) and derive the mean square error (MSE) of MLE to measure the quality of an estimator. We then define DM-optimality criterion, which is based on a function of the MSE. This criterion is applied to obtain locally optimal designs for small size experiments. The effect of sample size on the obtained designs are shown. We also obtain locally DM-optimal designs for some special cases of the model.

Keywords: bias, DM-optimal criterion, mean square error, Poisson regression model

1. Introduction

The validity of an experiment depends on the appropriate experimental design that can be elucidated by using the theory of optimum experimental design. A convenient framework to provide the ideal design is to select a set of control variables so that it leads to the minimum of the variance of parameter estimator. There is an increasing number of experiments that a linear model cannot adequately indicate the fundamental features of the data.

Observational studies involve many situations in which the outcome of an experiment is countable. An appropriate model to describe such data is the Poisson regression model, that can be found in McCullagh and Nelder (1989). Unfortunately, there are difficulties with the optimal design problem in such models. These difficulties are caused by likelihood equations that are usually the nonlinear functions of the parameters. Therefore, the maximum likelihood estimator (MLE) cannot be solved analytically. Under such circumstances, numerical methods such as Fisher scoring are employed to obtain MLE (Searle *et al.*, 1992). Due to the lack of a closed form for MLE, variance-covariance matrix cannot be directly calculated. However, MLE has substantial asymptotic properties despite this limitation. According to the large-sample theory, MLE is asymptotically unbiased and the variance of MLE is approximated by the inverse of the Fisher information matrix. Optimal designs for Poisson regression models have been done by Wang *et al.* (2006a, b), Russell *et al.* (2009b). In these literature references, optimal designs are obtained based on the maximization of a convex function of the Fisher information matrix, which follow the asymptotic theory. An important question rising here is, whether

¹ Corresponding author: Razi University, Department of Statistics, University Avenue, Taghe Bostan, Kermanshah, Iran.
E-mail: Niaparast@razi.ac.ir

the asymptotic properties establish when the sample size is not large enough. Numerical results show that the asymptotic approximations do not hold for small size experiments. Therefore, the mentioned problem may affect the accuracy of the obtained optimal designs.

For non-linear models, the small-sample differential-geometric approach based on the least squares estimator is deemed to improve the experiment conclusions (Pronzato and Pazman, 2013). Russell *et al.* (2009a) considered a locally optimal design for the simple logistic regression model in small size experiments. They used the maximum penalized likelihood, which has been studied by Firth (1993), for estimation of parameters and obtained locally optimal designs for two support points of the designs. Poursina and Talebi (2013) proposed a modified D-optimality criterion based on the Bhattacharyya lower bound for two parameters simple logistic model. They presented an intuitive proof and demonstrated that modified optimal designs are more efficient than the previous optimal designs. Khuri *et al.* (2006) derived the mean square error (MSE) for the linear predictor, and consider optimal design for the generalized linear models (GLMs). Mehr Mansour and Niaparast (2019) have also shown a numerical study on optimal designs for small sample size in logistic regression model.

In the present work, we provide the discussion of the higher-order asymptotic expansion of the bias of the MLE for the Poisson regression model. In order to obtain accurate results in small size experiments, it is preferable that the variance and the bias are considered together to obtain optimal designs (Box and Draper, 1959). To fulfil this, we primarily introduce the Poisson regression model and review the asymptotic expansion of MLE for this model. We then discuss the bias of MLE for the Poisson regression model and obtain the MSE by using the second-order expansion of MLE. In Section 3, we define DM-optimality criterion, and then study optimal designs for the special cases of the Poisson regression model based on the minimization of the MSE.

2. Poisson regression model

The main assumption in the linear regression is that the response variable is normally distributed. There are numerous conditions in which outcomes have non-normal distribution, but the random response is a member of the exponential family. Nelder and Wedderburn (1972) established the GLMs to describe such data. The Poisson regression model is a special case of GLMs used to model count data where the response variable is specified by the Poisson distribution. Some applications of this model in medicine, social sciences and toxicology have been studied, for example see Coxe *et al.* (2009) and Cameron and Trivedi (2013).

Suppose that Y_i 's are independent random variables from Poisson distribution which assume that the logarithm of their means can be modelled by a linear combination of unknown parameters, i.e.

$$\log(\mu_i) = \mathbf{f}^T(\mathbf{x}_i)\boldsymbol{\beta}, \quad i = 1, \dots, n,$$

where $\boldsymbol{\beta}^T = (\beta_1, \dots, \beta_p)$ is the vector of unknown parameters, $\mu_i = \mu_i(\mathbf{x}_i, \boldsymbol{\beta})$ is the mean of Y_i and $\mathbf{f}^T(\mathbf{x}_i) = (f_1(\mathbf{x}_i), \dots, f_p(\mathbf{x}_i))$ is the vector of known regression functions at explanatory variables, \mathbf{x}_i , that are the same for all individuals.

This model has been widely studied in literature, for example see McCullagh and Nelder (1989) and Cameron and Trivedi (2013).

2.1. Maximum likelihood estimator

For the Poisson regression model, the log likelihood can be written as

$$L(\boldsymbol{\beta}) = \sum_{i=1}^n \left[y_i \mathbf{f}^T(\mathbf{x}_i) \boldsymbol{\beta} - e^{\mathbf{f}^T(\mathbf{x}_i) \boldsymbol{\beta}} - \log(y_i!) \right],$$

and

$$\mathbf{l} = \frac{\partial L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^n \left[y_i \mathbf{f}(\mathbf{x}_i) - \mathbf{f}(\mathbf{x}_i) e^{\mathbf{f}^T(\mathbf{x}_i) \boldsymbol{\beta}} \right].$$

The score function to estimate the Poisson regression parameters is completely non-linear; therefore, the MLE of parameters cannot be obtained in a closed form. This problem causes a difficulty for making inference based on the maximum likelihood estimation. The common approach is to use the asymptotic approximation of the score function, the derivative of the log-likelihood function, that indicates sensitive a likelihood function with respect to parameters.

To denote the different orders of derivatives of the log-likelihood function, we apply the following notations.

$$\begin{aligned} l_r &= \frac{\partial L(\boldsymbol{\beta})}{\partial \beta_r} = \sum_{i=1}^n \left(-f_r(\mathbf{x}_i) e^{\mathbf{f}^T(\mathbf{x}_i) \boldsymbol{\beta}} + y_i f_r(\mathbf{x}_i) \right), \\ L_{rs} &= \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_r \partial \beta_s} = \sum_{i=1}^n \left(-f_r(\mathbf{x}_i) f_s(\mathbf{x}_i) e^{\mathbf{f}^T(\mathbf{x}_i) \boldsymbol{\beta}} \right), \\ L_{rst} &= \frac{\partial^3 L(\boldsymbol{\beta})}{\partial \beta_r \partial \beta_s \partial \beta_t} = \sum_{i=1}^n \left(-f_r(\mathbf{x}_i) f_s(\mathbf{x}_i) f_t(\mathbf{x}_i) e^{\mathbf{f}^T(\mathbf{x}_i) \boldsymbol{\beta}} \right). \end{aligned}$$

We also indicate the expectation of the above expressions with

$$\lambda_{rs} = E(L_{rs}), \quad \lambda_{rst} = E(L_{rst}).$$

The Taylor expansion of the r^{th} element of the score function, $\hat{l}_r = l_r(\hat{\boldsymbol{\beta}})$, about $\boldsymbol{\beta}$ is

$$\hat{l}_r = l_r + \sum_s L_{rs} (\hat{\beta}_s - \beta_s) + \frac{1}{2} \sum_{s,t} L_{rst} (\hat{\beta}_s - \beta_s) (\hat{\beta}_t - \beta_t) + \dots \quad (2.1)$$

Then Equation (2.1) is inverted to obtain an asymptotic expansion for $(\hat{\beta}_r - \beta_r)$. Therefore, asymptotic expansion for $(\hat{\beta}_r - \beta_r)$ can be written as

$$(\hat{\beta}_r - \beta_r) = - \sum_s L^{rs} l_s - \frac{1}{2} \sum_{s,t} a_{rst} l_s l_t + \sum_{s,t,u} L^{rs} L^{tu} H_{st} l_u + O_P(n^{-\frac{3}{2}}), \quad (2.2)$$

where, $a_{rst} = \sum_{g,h,i} L^{rg} L^{sh} L^{ti} L_{ghi}$, $H_{st} = L_{st} - \lambda_{st}$ and $\{L^{rs}\}_{r,s=1}^p$ is the inverse matrix of $\{L_{rs}\}_{r,s=1}^p$ (e.g., for example Lawley (1956)).

The bias of $\hat{\beta}_r$, $r = 1, \dots, p$, can be represented as follow,

$$\text{bias}(\hat{\beta}_r) = \frac{1}{2} \sum_{i,j,h} i^{ri} j^{jh} \lambda_{ijh}, \quad (2.3)$$

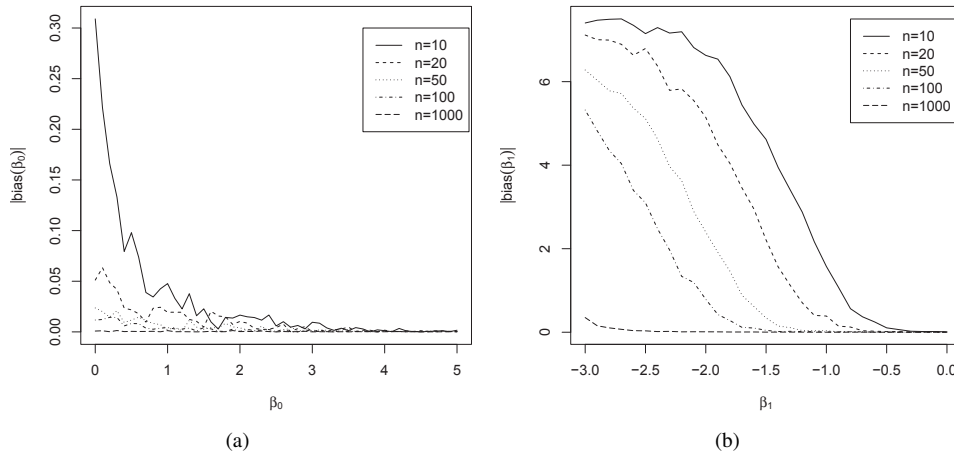


Figure 1: (a): $|\text{bias}(\hat{\beta}_0)|$ when $\beta_1 = -1$; (b): $|\text{bias}(\hat{\beta}_1)|$ when $\beta_0 = 1$.

where, $\{i^{rs}\}_{r,s=1}^p$ is the inverse matrix of the Fisher information matrix, $\mathbf{I}(\boldsymbol{\beta}) = \{-\lambda_{rs}\}_{r,s=1}^p$ (see for example Barndorff-Nielsen and Cox (1994, p.150)).

It can be obtained from Equation (2.2) that the above bias, called the second order bias, is of order $O(n^{-3/2})$. Whereas according to the large sample theory, the first order bias, which is based on the first term of right side of Equation (2.2), tends to zero (Lehmann, 1999).

To illustrate the influence of the sample size on the bias of $\boldsymbol{\beta}$, we now consider the following simulated example.

Example 1. Consider a simple Poisson regression model, where the response mean is described in terms of an explanatory variable

$$\mu_i = e^{\beta_0 + \beta_1 x_i}, \quad i = 1, \dots, n.$$

Furthermore, let x_i 's take values $\{0, 2\}$. For the purpose of illustration, we have simulated a data set of the simple Poisson regression model, so that half of the observations is taken at $x = 0$ and the other half at $x = 2$. The number of observations is set in sizes $n = 10, 20, 50, 100, 1000$. The number of simulations is $s = 1000$ in each case.

In each replication, we fit the simple Poisson model and use the simulated data and the explanatory variables to compute the average bias of the parameter estimates for each combination of $(n, \boldsymbol{\beta})$ defined as

$$\text{bias}(\hat{\beta}_r) = \frac{1}{s} \sum_{j=1}^s (\hat{\beta}_r^j - \beta_r),$$

where $\hat{\beta}_r^j$ is the estimator of β_r for j^{th} replication. Figure 1 shows the simulated results.

Figure 1(a) reports the absolute value of $\text{bias}(\hat{\beta}_0)$ as a function of β_0 with the assumption that $\beta_1 = -1$. For any fixed value of β_0 , the absolute of $\text{bias}(\hat{\beta}_0)$ is a decreasing function of n . The plots in Figure 1(a) demonstrate that the absolute of $\text{bias}(\hat{\beta}_0)$ tends to zero when β_0 is rising.

Figure 1(b) shows the absolute value of the bias of $\hat{\beta}_1$ as a function of β_1 , when $\beta_0 = 1$. Similar to Figure 1(a), it is observed that the absolute value of $\text{bias}(\hat{\beta}_1)$ is a decreasing function of n specially for small value of β_1 .

Table 1: The bias for the simple Poisson model

		n				
		10	16	20	30	50
bias ($\hat{\beta}_0$)	Simulated bias	-0.036	-0.023	-0.021	-0.001	-0.009
	Second order bias	-0.037	-0.023	-0.018	-0.012	-0.007
bias ($\hat{\beta}_1$)	Simulated bias	-1.481	-0.536	-0.291	-0.072	-0.025
	Second order bias	-0.118	-0.073	-0.059	-0.039	-0.024

As a general conclusion, the simulated results have indicated that the absolute value of the bias of MLE of parameters is a decreasing function of distribution means. It is also shown that the bias of MLE is not zero in small sample experiments. While, the first order bias is equal to zero. Therefore, it is useful to compare the simulated bias with the second order bias which is expressed in Equation (2.3). Note that the true values of the parameters are $\beta_0 = 1$ and $\beta_1 = -1$.

Table 1 shows that the second order bias is closer to the simulated bias and is a better approximation for computation bias. Therefore, we apply Equation (2.2) to inference about MLE.

2.2. Mean square error

The MSE is a useful tool to measure the difference between the estimator and the true value of the parameters. The MSE expresses the quality of an estimator, $\hat{\beta}$, and is defined as follows,

$$\begin{aligned} \text{MSE}(\hat{\beta}) &= E(\hat{\beta} - \beta)(\hat{\beta} - \beta)^T \\ &= \text{Var}(\hat{\beta}) + \text{bias}(\hat{\beta})\text{bias}^T(\hat{\beta}), \end{aligned}$$

where $\text{MSE}(\hat{\beta})$ is a symmetric and positive semi-definite matrix. The $\text{MSE}(\hat{\beta})$ incorporates both the variance of the estimator and its bias.

Note that if we apply the first order of the Taylor expansion of the score function, Equation (2.1), or we ignore the second term of right side of Equation (2.2), then the bias of β tends to zero. The $\text{MSE}(\hat{\beta})$ will be equal to $\mathbf{I}^{-1}(\beta)$ asymptotically.

In the following theorem, we obtain $\text{MSE}(\hat{\beta})$ for the Poisson regression model.

Theorem 1. Consider a Poisson regression model with canonical link function. The MSE for the MLE of the vector of parameters can be represented as

$$\begin{aligned} \text{MSE}(\hat{\beta}) &\simeq \mathbf{I}^{-1}(\beta) - \frac{1}{2}\mathbf{I}^{-1}(\beta)\mathbf{F}^T\mathbf{W}\mathbf{U}^{(2)}\mathbf{W}\mathbf{F}\mathbf{I}^{-1}(\beta) \\ &\quad + \frac{1}{4}\mathbf{I}^{-1}(\beta)\mathbf{F}^T\mathbf{W}\mathbf{U}^{(2)}\mathbf{W}\mathbf{U}^{(2)}\mathbf{W}\mathbf{F}\mathbf{I}^{-1}(\beta) \\ &\quad + \frac{1}{4}\mathbf{I}^{-1}(\beta)\mathbf{F}^T\mathbf{W}\mathbf{D}_U\mathbf{J}\mathbf{D}_U\mathbf{W}\mathbf{F}\mathbf{I}^{-1}(\beta), \end{aligned} \quad (2.4)$$

where $\mathbf{W} = \text{diag}\{\mu_i\}_{i=1}^n$ is a diagonal matrix of means, $\mathbf{F} = \{\mathbf{f}(\mathbf{x}_1), \dots, \mathbf{f}(\mathbf{x}_n)\}^T$ is the design matrix, $\mathbf{I}(\beta) = \mathbf{F}^T\mathbf{W}\mathbf{F}$, $\mathbf{U} = \mathbf{F}\mathbf{I}^{-1}(\beta)\mathbf{F}^T$, $\mathbf{U}^{(2)} = \{u_{ij}^2\}_{i,j=1}^n$, $\mathbf{D}_U = \text{diag}\{u_{11}, \dots, u_{nn}\}$ and \mathbf{J} is a $n \times n$ matrix of unit elements.

The proof of Theorem 1 is given in Appendix A.

Khuri *et al.* (2006) derive MSE for the linear predictor using the asymptotic variance and the second order bias, while we directly obtain the MSE of the second order expansions of the MLE for the vector estimators.

3. DM-optimal design

Optimal design is to provide the best estimation of parameters based on specific design criterion. Most optimal criteria are based on the variance of the estimator of parameters. For the Poisson regression model, Wang *et al.* (2006a, b) conducted extensive work on optimal design based on a convex function of the Fisher information matrix. However, for small size experiments, the Fisher information matrix may not be a trustworthy approximation to obtain the optimal designs. Therefore, we propose to use MSE to obtain optimal design for small size experiments due better accuracy. Suppose that $\xi = \begin{Bmatrix} x_1 & \cdots & x_r \\ n_1 & \cdots & n_r \end{Bmatrix}$ is an exact design where x_i 's are points of design and n_i is the number of observations which are taken at x_i , $\sum_{i=1}^r n_i = n$. The design ξ^* is better than the design ξ , if $MSE(\hat{\beta}_\xi) - MSE(\hat{\beta}_{\xi^*})$ is a non-negative definite matrix. Unfortunately, it is not possible to solve the above optimization problem. Therefore, the design of experiments is selected so that some real functions of the MSE of estimators are minimized. The most popular approach is based on the determinant of the $MSE(\hat{\beta}_\xi)$. Therefore, we define a new optimality criterion as follows,

Definition 1. A design ξ^* is said to be DM-optimal in the class of designs Ξ if

$$\log \det \left(MSE(\hat{\beta}_{\xi^*}) \right) \leq \log \det \left(MSE(\hat{\beta}_\xi) \right) \quad \forall \xi \in \Xi, \quad (3.1)$$

where D and M are abbreviation of the determinant and the MSE.

This criterion is different of the D-optimality criterion. DM-optimality criterion is based on the determinant of the MSE whereas D-optimality criterion is based on the determinant of the Fisher information matrix.

There are two essential problems: First, $MSE(\hat{\beta}_\xi)$ depends on the unknown parameters through response mean, and therefore, the DM-optimality criterion depends on the vector of parameters, β . Whereas the exact values of β are unknown in practice. We consider the locally DM-optimal design. Second, $MSE(\hat{\beta}_\xi)$ relies on the sample size. It means that the sample size affects the DM-optimal design. $MSE(\hat{\beta}_\xi)$ is based on the more accurate approximation of MLE; therefore, it is expected that DM-optimal designs are more efficient for small size experiments.

For any ξ , there exists a value of the $\det(MSE(\hat{\beta}_\xi))$. We need an optimization method to find the design ξ^* that minimize $\log \det(MSE(\hat{\beta}_\xi))$. Here, we have a discrete optimization problem which is difficult to find a solution for this problem. An approach is to consider all possible combinations of n_i 's to obtain DM-optimal design. We can use also approximate designs to guess the optimum number of trials at any point of design instead of this calculation. Approximate designs ignore the constraint that the number of observations must be an integer at any point of design. These results are then employed to find DM-optimal designs based on the number of trials at any point of design.

Optimization classical methods often are known as local optimization since they derive the different results with different starting points. Therefore in this paper, we apply a hybrid method which uses of two optimality algorithm to obtain the exact global optimal point.

First a genetic algorithm (GA) is used to find near point to global optimal point (Chen and Ye, 2009). The GA is a global optimization algorithm based on the principles of natural selection and genetics. This process starts with a population of individuals, which serve as the first generation to solve above optimization problem. Then selection, crossover and mutation operations are employed to produce the next generation. It is repeated until a sufficiently large number of generations have passed without any improvement in the value of the function that we want to optimize. We use the

second optimization due to the constraints of GA that, do not guarantee the exact global optimum despite resulting in a near global optimum.

Second, a local optimality algorithm is employed to achieve the global optimal point using the obtained point in GA as the starting point. We use of the sequential quadratic programming (SQP) method work out in “fmincon” function at MATLAB program (Fletcher and Powell, 1963).

In this paper, we consider two special cases of the Poisson regression model. Equations (3.2) and (3.3) define the mean function for simple model and quadratic model, respectively,

$$\mu_i = \exp(\beta_0 + \beta_1 x_i), \tag{3.2}$$

$$\mu_i = \exp(\beta_0 + \beta_1 x_i + \beta_2 x_i^2). \tag{3.3}$$

In most applications of this model, especially in toxicity studies, it is assumed that the design region is the non-negative subset of real numbers, $\mathcal{X} = [h, g]$ where h is a non-negative real number and $g > h$ (Wang *et al.*, 2006a). In this study, we suppose that $\mathcal{X} = [0, \infty]$. Without loss of generality, we consider a special case, where μ_i is assumed to be a decreasing function of x_i . We define canonical mean $\tilde{\mu}_i = \mu_i / \mu_{\max}$, where μ_{\max} is the maximum value of $\mu_i = \mu_i(\mathbf{x}_i, \boldsymbol{\beta})$, which is achieved in the lower bound of design region. It is clear that $\tilde{\mu}_i$ will be in $[0, 1]$. In this paper, we adopt a design space in terms of $\tilde{\mu}_i$'s rather than x_i 's. According to the monotone relation between $\tilde{\mu}_i$ and x_i , we can consider designs based on both, equivalently. All the designs in this paper are derived under this assumption; however, they can be easily generalized to situations that μ_i is an increasing function of x_i .

3.1. Locally DM-optimal design for the simple model

In this section, we consider a simple Poisson regression model with mean μ_i , which has been specified in (3.2). We are also restricted in design space, Ξ , to saturated designs or designs with two support points, i.e., $\xi = \left\{ \begin{matrix} \tilde{\mu}_1 & \tilde{\mu}_2 \\ n_1 & n_2 \end{matrix} \right\}$. Note that $\tilde{\mu}_i = e^{\beta_1 x_i}$ or equivalently $x_i = (1/\beta_1) \log(\tilde{\mu}_i)$. The following theorem obtains $MSE(\hat{\boldsymbol{\beta}}_\xi)$ for this model.

Theorem 2. For a simple Poisson regression model, $MSE(\hat{\boldsymbol{\beta}}_\xi)$ can be represented as

$$MSE(\hat{\boldsymbol{\beta}}_\xi) \approx \frac{1}{a} \begin{pmatrix} m_{11} & m_{12} \\ m_{12} & m_{22} \end{pmatrix},$$

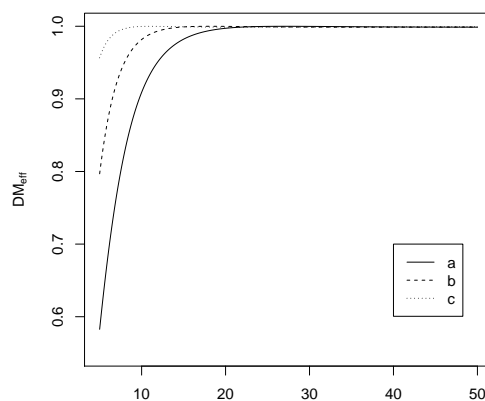
where $a = (\log(\tilde{\mu}_1) - \log(\tilde{\mu}_2))^2 e^{3\beta_0} n_1^3 n_2^3 \tilde{\mu}_1^3 \tilde{\mu}_2^3$,

$$\begin{aligned} m_{11} &= e^{2\beta_0} n_1^2 n_2^2 \tilde{\mu}_1^2 \tilde{\mu}_2^2 \sum n_i \tilde{\mu}_i \log^2(\tilde{\mu}_i) - 0.25 e^{\beta_0} n_1 n_2 \tilde{\mu}_1 \tilde{\mu}_2 \sum n_i^2 \tilde{\mu}_i^2 \log^2(\tilde{\mu}_i) \\ &\quad + 0.25 \sum n_i^3 \tilde{\mu}_i^3 \log^2(\tilde{\mu}_i) - 0.5 e^{\beta_0} n_1^2 n_2^2 \tilde{\mu}_1^2 \tilde{\mu}_2^2 \log(\tilde{\mu}_1) \log(\tilde{\mu}_2), \\ m_{12} &= -\beta_1 e^{2\beta_0} n_1^2 n_2^2 \tilde{\mu}_1^2 \tilde{\mu}_2^2 \sum n_i \tilde{\mu}_i \log(\tilde{\mu}_i) + 0.25 \beta_1 e^{\beta_0} n_1 n_2 \tilde{\mu}_1 \tilde{\mu}_2 \sum n_i^2 \tilde{\mu}_i^2 \log(\tilde{\mu}_i) \\ &\quad - 0.25 \beta_1 \sum n_i^3 \tilde{\mu}_i^3 \log(\tilde{\mu}_i) + 0.25 e^{\beta_0} n_1^2 n_2^2 \tilde{\mu}_1^2 \tilde{\mu}_2^2 (\log(\tilde{\mu}_1) + \log(\tilde{\mu}_2)), \\ m_{22} &= \beta_1^2 e^{2\beta_0} n_1^2 n_2^2 \tilde{\mu}_1^2 \tilde{\mu}_2^2 \sum n_i \tilde{\mu}_i - 0.25 \beta_1^2 e^{\beta_0} n_1 n_2 \tilde{\mu}_1 \tilde{\mu}_2 \sum n_i^2 \tilde{\mu}_i^2 \\ &\quad + 0.25 \beta_1^2 \sum n_i^3 \tilde{\mu}_i^3 - 0.5 \beta_1^2 e^{\beta_0} n_1^2 n_2^2 \tilde{\mu}_1^2 \tilde{\mu}_2^2. \end{aligned}$$

Proof: For any $\xi = \left\{ \begin{matrix} \tilde{\mu}_1 & \tilde{\mu}_2 \\ n_1 & n_2 \end{matrix} \right\}$, we have $\mathbf{F}_\xi = \begin{pmatrix} 1 & \frac{1}{\beta_1} \log(\tilde{\mu}_1) \\ 1 & \frac{1}{\beta_1} \log(\tilde{\mu}_2) \end{pmatrix}$ and $\mathbf{W}_\xi = \begin{pmatrix} n_1 \mu_1 & 0 \\ 0 & n_2 \end{pmatrix}$ which are design matrix

Table 2: DM-optimal design for the simple model, $n_2 = n - n_1$

n	$\beta_0 = 0.1, \beta_1 = -1$			$\beta_0 = 0.5, \beta_1 = -1$			$\beta_0 = 1, \beta_1 = -1$		
	n_1	$\tilde{\mu}_1$	$\tilde{\mu}_2$	n_1	$\tilde{\mu}_1$	$\tilde{\mu}_2$	n_1	$\tilde{\mu}_1$	$\tilde{\mu}_2$
8	5	0.1705	1	4	0.1669	1	4	0.1402	1
10	6	0.1607	1	5	0.1532	1	5	0.1328	1
15	8	0.1483	1	8	0.1388	1	8	0.1270	1
20	10	0.1396	1	10	0.1286	1	10	0.1249	1
30	15	0.1285	1	15	0.1250	1	15	0.1262	1
50	25	0.1249	1	25	0.1262	1	25	0.1290	1

Figure 2: DM_{eff} for DM-optimal design in the simple model: (a) $\beta_0 = 0.1$; (b) $\beta_0 = 0.5$; (c) $\beta_0 = 1$.

and diagonal matrix of means corresponding to design ξ , respectively. Then

$$\mathbf{I}(\boldsymbol{\beta}) = \frac{e^{\beta_0}}{\beta_1^2} \begin{pmatrix} \beta_1^2 \sum n_i \tilde{\mu}_i & \beta_1 \sum n_i \tilde{\mu}_i \log(\tilde{\mu}_i) \\ \beta_1 \sum n_i \tilde{\mu}_i \log(\tilde{\mu}_i) & \sum n_i \tilde{\mu}_i \log^2(\tilde{\mu}_i) \end{pmatrix},$$

$$\mathbf{U} = \frac{1}{e^{\beta_0}} \begin{pmatrix} \frac{1}{n_1 \tilde{\mu}_1} & 0 \\ 0 & \frac{1}{n_2 \tilde{\mu}_2} \end{pmatrix}.$$

Replacing the above matrices in Equation (2.4), and then straightforward calculation leads to the results. \square

We have assumed that μ_i is a decreasing function of x_i . $x_i \in [0, \infty]$; therefore, this implies that $\beta_1 < 0$ for simple model in (3.2). Using the Theorem 2 and applying hybrid method, the results for DM-optimal design are obtained for some representative values of parameters in Table 2.

DM-optimal design optimizes the determinant of $\text{MSE}(\hat{\boldsymbol{\beta}}_{\xi^*})$; therefore, it is beneficent that we compare the accuracy of DM-optimal design, ξ^* , with D-optimal design, ξ^+ , that maximizes the determinant of $\mathbf{I}(\boldsymbol{\beta})$. Therefore, we define DM_{eff} criterion as follows,

$$DM_{\text{eff}} = \frac{\det(\text{MSE}(\hat{\boldsymbol{\beta}}_{\xi^*}))}{\det(\text{MSE}(\hat{\boldsymbol{\beta}}_{\xi^+}))}. \quad (3.4)$$

Figure 2 shows DM_{eff} of obtained designs in Table 2.

Table 3: DM-optimal design for the quadratic model, $n_3 = n - n_1 - n_2$

n	$\beta_0 = 1, \beta_1 = -1, \beta_2 = -1$					$\beta_0 = 2, \beta_1 = -1, \beta_2 = -1$				
	n_1	n_2	$\tilde{\mu}_1$	$\tilde{\mu}_2$	$\tilde{\mu}_3$	n_1	n_2	$\tilde{\mu}_1$	$\tilde{\mu}_2$	$\tilde{\mu}_3$
8	4	2	0.0943	0.5416	1	3	2	0.0546	0.5079	1
10	4	3	0.0800	0.5303	1	4	3	0.0478	0.5017	1
15	6	4	0.0639	0.5163	1	5	5	0.0438	0.4983	1
20	8	6	0.0551	0.5088	1	7	6	0.0394	0.4934	1
30	10	10	0.0497	0.5044	1	10	10	0.0371	0.4911	1
50	17	16	0.0407	0.4948	1	16	17	0.0365	0.4907	1
n	$\beta_0 = 1, \beta_1 = -2, \beta_2 = -1$					$\beta_0 = 1, \beta_1 = -1, \beta_2 = -2$				
	n_1	n_2	$\tilde{\mu}_1$	$\tilde{\mu}_2$	$\tilde{\mu}_3$	n_1	n_2	$\tilde{\mu}_1$	$\tilde{\mu}_2$	$\tilde{\mu}_3$
8	4	2	0.0694	0.4571	1	3	2	0.0995	0.5760	1
10	4	3	0.0697	0.4576	1	4	3	0.0848	0.5652	1
15	6	4	0.0547	0.5163	1	5	5	0.0754	0.5560	1
20	8	6	0.0464	0.4337	1	7	6	0.0634	0.5488	1
30	10	10	0.0411	0.4281	1	10	10	0.0541	0.5409	1
50	17	16	0.0315	0.4155	1	16	17	0.0463	0.5336	1

The results show that DM-optimal designs are more efficient for small size experiments. DM-optimal designs are robust based on $\tilde{\mu}$; therefore, do not change for different values of β_1 . Figure 2 is drawn only for different values of β_0 . Plots for some different values of β_0 in Figure 2 show that the DM-optimal designs and the D-optimal designs will be very close to each other when β_0 increases.

3.2. Locally optimal design for the Quadratic model

Considering the effect of an explanatory variable on the response variable, the effect of explanatory variables is sometimes stronger than the simple model describes. Therefore, the quadratic model defined in (3.3) may be suitable. Here we survey DM-optimal design for the quadratic model. According to the decreasing assumption on the mean response, we take negative values for β_1 and β_2 . To find DM-optimal design, the class of designs, Ξ , has been restricted to the saturated designs with three support points, i.e., $\xi = \left\{ \begin{matrix} \tilde{\mu}_1 & \tilde{\mu}_2 & \tilde{\mu}_3 \\ n_1 & n_2 & n_3 \end{matrix} \right\}$ where, $\tilde{\mu}_i = e^{\beta_1 x_i + \beta_2 x_i^2}$ is equivalent to

$$x_i = \frac{1}{2\sqrt{-\beta_2}} \left[-\sqrt{-r} + \sqrt{-r - 4 \log(\tilde{\mu}_i)} \right],$$

where $r = \beta_1^2 / \beta_2$. Therefore, $\mathbf{F}_\xi = \begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \end{pmatrix}$ and $\mathbf{W}_\xi = \text{diag}\{n_i \mu_i\}_{i=1}^3$ are the design matrix and the diagonal matrix of means, respectively. DM-optimal designs have been obtained for some representative values of parameters (Table 3).

The DM_{eff} of DM-optimal designs for the quadratic model are illustrated in Figure 3. It is observed that DM-optimal designs are more accurate for small size experiments.

4. Conclusion

The non-existence of an explicit form for the parameter estimators of non-linear regression models leads statisticians to apply asymptotic inference based on the likelihood function to obtain the optimal design for GLMs.

In this study, we consider the Poisson regression model, which is a special case of the GLMs. The results indicate that for small size experiments, the asymptotic properties of parameter estimators; such as unbiasedness, may be in conflict. Therefore, we replace variance with MSE, which is consid-

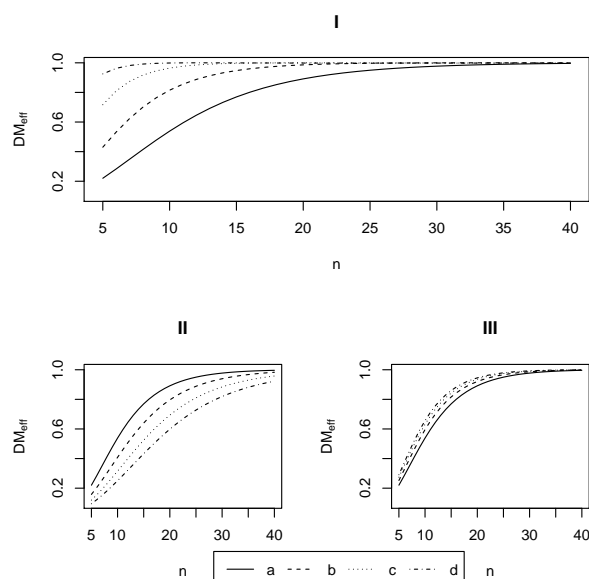


Figure 3: DM_{eff} for DM-optimal design in the quadratic model. I: $\beta_1 = -1, \beta_2 = -1$, (a) $\beta_0 = 1$; (b) $\beta_0 = 1.5$; (c) $\beta_0 = 2$; (d) $\beta_0 = 2.5$. II: $\beta_0 = 1, \beta_2 = -1$, (a) $\beta_1 = -1$; (b) $\beta_1 = -1.5$; (c) $\beta_1 = -2$; (d) $\beta_1 = -2.5$. III: $\beta_0 = 1, \beta_1 = -1$, (a) $\beta_2 = -1$; (b) $\beta_2 = -1.5$; (c) $\beta_2 = -2$; (d) $\beta_2 = -2.5$.

ered as both a variance and bias, to obtain optimal design. In addition, we obtained a more precise estimation of parameter β using the higher-order expansion of the likelihood function. To illustrate the theoretical results, we consider two special cases of the Poisson regression model. DM-optimal designs are calculated for small size experiments and compared with D-optimal designs, respectively. The results show that the optimal designs based on the MSE are better than the optimal designs based on the inverse of the information matrix.

An ongoing research activity is the investigation of Bayesian optimal designs instead of the local optimal designs.

Appendix:

Proof of Theorem 1: For Poisson regression model $(r, k)^{th}$ element of $MSE(\hat{\beta})$ is given by

$$\begin{aligned}
 m_{rk} &= E \left\{ (\hat{\beta}_r - \beta_r) (\hat{\beta}_k - \beta_k) \right\} \\
 &\simeq E \left(\sum_{s_1} i^{rs_1} l_{s_1} - \frac{1}{2} \sum_{s_1, t_1} \alpha_{rs_1 t_1} l_{s_1} l_{t_1} \right) \left(\sum_{s_2} i^{ks_2} l_{s_2} - \frac{1}{2} \sum_{s_2, t_2} \alpha_{ks_2 t_2} l_{s_2} l_{t_2} \right) \\
 &= \sum_{s_1, s_2} i^{rs_1} i^{ks_2} E(l_{s_1} l_{s_2}) + \frac{1}{4} \sum_{s_1, t_1, s_2, t_2} \alpha_{rs_1 t_1} \alpha_{ks_2 t_2} E(l_{s_1} l_{t_1} l_{s_2} l_{t_2}) \\
 &\quad - \frac{1}{2} \sum_{s_1, s_2, t_2} i^{rs_1} \alpha_{ks_2 t_2} E(l_{s_1} l_{s_2} l_{t_2}) - \frac{1}{2} \sum_{s_1, t_1, s_2} i^{ks_2} \alpha_{rs_1 t_1} E(l_{s_1} l_{t_1} l_{s_2}), \tag{A.1}
 \end{aligned}$$

where $\alpha_{rs_1 t_1} = - \sum_{q_1, q_2, q_3} i^{rq_1} i^{s_1 q_2} i^{t_1 q_3} \lambda_{q_1 q_2 q_3}$. Note that $H_{st} = 0$.

To obtain the expectation of the log-likelihood derivatives, we use a given general form by Barndorff-Nielsen and Cox (1994, p 146) that is as follow

$$\sum_{\pi=1}^R \sum_{R/\pi} v_{R_1, \dots, R_\pi} = 0, \quad (\text{A.2})$$

where $R = r_1 \cdots r_m$ is a set of coordinate indices. The inner sum in equation (A.2) is over all partitions of R into π blocks R_1, \dots, R_π and $v_{R_1, \dots, R_\pi} = E(L_{R_1} \cdots L_{R_\pi})$. Therefore, for the Poisson regression model, we obtain

$$\begin{aligned} E(l_{s_1} l_{s_2}) &= -\lambda_{s_1 s_2}, \\ E(l_{s_1} l_{s_2} l_{t_2}) &= -\lambda_{s_1 s_2 t_2}, \\ E(l_{s_1} l_{t_1} l_{s_2} l_{t_2}) &= \lambda_{s_1 t_1} \lambda_{s_2 t_2} + \lambda_{s_1 s_2} \lambda_{t_1 t_2} + \lambda_{s_1 t_2} \lambda_{s_2 t_1} - \lambda_{s_1 t_1 s_2 t_2}. \end{aligned}$$

Then, we have

$$\sum_{s_1, s_2} i^{r s_1} i^{k s_2} E(l_{s_1} l_{s_2}) = i^{r k}.$$

Moreover, we can find

$$\begin{aligned} & \sum_{s_1, s_2, t_2} i^{r s_1} \alpha_{k s_2 t_2} E(l_{s_1} l_{s_2} l_{t_2}) \\ &= \sum_{i, m} \mu_i \mu_m \left(\sum_{s_1} f_{s_1}(\mathbf{x}_m) i^{r s_1} \right) \left(\sum_{s_2, q_2} f_{s_2}(\mathbf{x}_i) f_{q_2}(\mathbf{x}_m) i^{s_2 q_2} \right) \left(\sum_{t_2, q_3} f_{t_2}(\mathbf{x}_m) f_{q_3}(\mathbf{x}_i) i^{t_2 q_3} \right) \left(\sum_{q_1} f_{q_1}(\mathbf{x}_i) i^{k q_1} \right) \\ &= \sum_{i, m} \mu_i \mu_m u_{im}^2 \left(\sum_{s_1} f_{s_1}(\mathbf{x}_m) i^{r s_1} \right) \left(\sum_{q_1} f_{q_1}(\mathbf{x}_i) i^{k q_1} \right). \end{aligned}$$

Similarly,

$$\sum_{s_1, t_1, s_2} i^{k s_2} \alpha_{r s_1 t_1} E(l_{s_1} l_{t_1} l_{s_2}) = \sum_{i, m} \mu_i \mu_m u_{im}^2 \left(\sum_{s_2} f_{s_2}(\mathbf{x}_m) i^{k s_2} \right) \left(\sum_{q_1} f_{q_1}(\mathbf{x}_i) i^{r q_1} \right).$$

We also get

$$\begin{aligned} \sum_{s_1, t_1, s_2, t_2} \alpha_{r s_1 t_1} \alpha_{k s_2 t_2} E(l_{s_1} l_{t_1} l_{s_2} l_{t_2}) &= 2 \sum_{i, m} \mu_i \mu_m u_{im}^2 \left(\sum_{q_1} f_{q_1}(\mathbf{x}_m) i^{r q_1} \right) \left(\sum_{p_1} f_{p_1}(\mathbf{x}_i) i^{k p_1} \right) \\ &+ \sum_{i, m} \mu_i \mu_m u_{ii} u_{mm} \left(\sum_{q_1} f_{q_1}(\mathbf{x}_i) i^{r q_1} \right) \left(\sum_{p_1} f_{p_1}(\mathbf{x}_m) i^{k p_1} \right) \\ &+ \sum_{i, j, m} \mu_i \mu_j \mu_m u_{ij}^2 u_{mj}^2 \left(\sum_{q_1} f_{q_1}(\mathbf{x}_i) i^{r q_1} \right) \left(\sum_{p_1} f_{p_1}(\mathbf{x}_m) i^{k p_1} \right). \end{aligned}$$

The results is obtained by substituting the above three equations in (A.1) and the straightforward application of matrix algebra. \square

References

- Barndorff-Nielsen OE and Cox DR (1994). *Inference and Asymptotics*, Chapman and Hall, London.
- Box GEP and Draper NR (1959). A basis for the selection of a response surface design, *Journal of the American Statistical Association*, **54**, 622–654.
- Cameron AC and Trivedi PK (2013). *Regression Analysis of Count Data*, Cambridge University Press, New York.
- Chen Y and Ye K (2009). Bayesian hierarchical modeling on dual response surfaces in partially replicated designs, *Quality Technology and Quantitative Management*, **6**, 371–389.
- Coxe S, West SG, and Aiken LS (2009). The analysis of count data: a gentle introduction to Poisson regression and its alternatives, *Journal of Personality Assessment*, **91**, 121–136.
- Firth D (1993). Bias reduction of maximum likelihood estimates, *Biometrika*, **80**, 27–38.
- Fletcher R and Powell MJD (1963). A rapidly convergent descent method for minimization, *Computer Journal*, **6**, 163–168.
- Khuri AI, Mukherjee B, Sinha BK, and Gosh M (2006). Design issues for generalized linear models: a review, *Statistical Science*, **21**, 376–399.
- Lawley DN (1956). A general method for approximating to the distribution of likelihood ratio criteria, *Biometrika*, **43**, 295–303.
- Lehmann EL (1999). *Elements of Large Sample Theory*, Springer-Verlag, New York.
- McCullagh P and Nelder JA (1989). *Generalized Linear Models* (2nd ed), Chapman and Hall, London.
- Mehr Mansour S and Niaparast M (2019). The effect of small sample on optimal designs for logistic regression models, *Communications in Statistics-Theory and Methods*, **48**, 2893–2903.
- Nelder JA and Wedderburn RWM (1972). Generalized linear models, *Journal of the Royal Statistical Society, Series A*, **135**, 370–384.
- Poursina D and Talebi H (2013). Modified D-optimal design for logistic model, *Journal of Statistical Computation and Simulation*, **84**, 428–437.
- Pronzato L and Pazman A (2013). *Design of Experiments in Nonlinear Models: Asymptotic Normality, Optimality Criteria and Small-Sample Properties*, volume 212 of *Lecture Notes in Statistics*, Springer, New York.
- Russell KG, Eccleston JA, Lewis SM, and Woods DC (2009a). Design considerations for small experiments and simple logistic regression, *Journal of Statistical Computation and Simulation*, **79**, 81–91.
- Russell KG, Woods DC, Lewis SM, and Eccleston JA (2009b). D-optimal designs for Poisson regression models, *Statistica Sinica*, **19**, 721–730.
- Searle SR, Casella G, and McCulloch CE (1992). *Variance Components*, Wiley, New York.
- Wang Y, Myers RH, Smith EP, and Ye K (2006a). D-optimal designs for Poisson regression models, *Journal of Statistical Planning Inference*, **136**, 2831–2845.
- Wang Y, Smith EP, and Ye K (2006b). Sequential designs for a Poisson regression model in toxicological and medical studies, *Journal of Statistical Planning Inference*, **136**, 3187–3202.