Stochastic Modeling and Applications
Vol.26 No. 1 (January-June, 2022)
ISSN: 0972-3641

*Received: 30th January 2022* *Revised: 05th March 2022* *Selected: 09th April 2022*

# D OPTIMAL DESIGN FOR POISSON MODEL

S. JOSHUA DAVID AND SWETHA. N*

Abstract. (One of the most widely used design criterion for linear regression models is D- Optimality. Binary or count data, such as defective/non-defective or number of flaws, often appear in industrial tests. Such data (GLMS) are suitable. Generalized Linear Models are particularly helpful in actuarial applications because they allow for the estimation of multiplicative models as well as types of heteroscedasticity such as Poisson-type and gamma-type with a fixed coefficient of variation, which are common in actuarial issues. The asymptotic covariance matrix, which is a weighted version of the covariance matrix for the linear case and an extension of the existing D-optimality technique, can be used to build a comparable D-optimality design criterion for GLM. Under a Poisson regression model with any number of independent variables and a square root link model linear predictor, we explore the problem of finding an optimal design. To establish local D-optimality of a class of designs, a canonical form of the problem and a generalized equivalence theorem are utilised. The theorem is combined with clustering techniques to produce a quick method for locating designs that are resistant to a wide range of model parameter values.

## 1. Introduction

D-optimality is a typical linear regression model design, partly because it relates to minimising the area of the confidence region for unknown parameters, and partly because it is theoretically and computationally simple. There are three major components to consider in the theory of optimal designs. The statistical model that links the response (observation) to the explanatory variables (factors), the experimental region that represents the range of these factors, and the optimality criterion for the proposed model on an experimental region. The variance-covariance matrix of parameter estimates, or its inverse, the Fisher information matrix, is used to evaluate the design' s quality. The values of the model parameters that appear in the information matrix affect the solution of optimal designs for generalized linear models, making it difficult. The expected information matrix for GLMs is very simple, consisting of a weighted version of the linear model' s covariance matrix. As a result, we show how traditional D-optimal design techniques can be easily applied to GLMs.

In recent years, the field of biomedical and clinical trials has paid more attention to optimal experimental designs for Poisson regression models. The dependence of

***Key words and phrases.*** Design of experiments, Poisson model, clustering, locally optimum design, Log link models.
*Corresponding Author.

design support points on unknown parameters of the Fisher information matrix is a significant difficulty in the building and evolution of design for generic nonlinear models (Poisson included) (Zhang, 2006). The best designs for generalised linear models can't be found without knowing the parameters first (Khuri et al. (2006), Atkinson and Woods (2015)). Chernoff (1953) introduced a method called local optimality that aimed to derive a locally optimal design at a given parameter value. For example, see Wang et al. (2006), Russell et al. (2009), and Grabhoff, Holling, and Schwabe for count data with poisson models and the Rasch poisson model (2013,2015,2018). Locally optimal designs (kalish and Rosenberger, 1978; Myers et al.,1994),minimax designs (sitter,1992), Bayesian designs (Chaloner and Larntz, 1989) and multi stage designs (Mikin ,1987; Myers et al.,1996; Sitter and Wu,1999).

## 2. The D-Optimality Criterion for Poisson Model

Simple optimizations based on an optimality criterion and the model to be fitted are called D-optimal designs. Maximizing $|X'X|$, the determinant of the information matrix $X'X$, is the optimality criterion for constructing D-optimal designs. Using this optimality criterion, the generalized variance of parameter estimations for a pre-specified model is minimized. As a result, a D-optimal design's 'optimality' is model dependent. Let Y1, Y2,..., Yn be independent Poisson-distributed response variables for n experimental units, with each Yi's density written as

$$\frac{\lambda_i^x}{x!} e^{-\lambda_i}, \quad \lambda_i > 0$$

Then, using the given "square root link" function g, the expected value of response Yi is related to the predictors Xi:

$$E\left[Y_i/x_i\right] = \mu_i = g(x_i, \beta) = g(\eta)$$

Where,

$$\mu = \eta_i^2$$

$\beta$ is a set of parameters that must be estimated. The variance of the response is calculated as follows

$$Var\left[Y_i/x_i\right] = V(\mu_i)$$

Beyond the mean and variance relations stated above, the expected information is independent of the distribution's form. According to Dobson (1983, Appendix 2), the element of the expected information matrix for GLMs with the above specification is as follows:

$$I_{jk} = \sum_i \frac{x_{ij} x_{jk}}{V(\mu_i)} \left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2$$

In matrix notation this would be

$$I = X' WX$$

Where W is a diagonal matrix having values

$$W_i = \frac{1}{V(\mu_i)} \left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2$$

As a result, the asymptotic variance covariance matrix

$$I^{-1} = (X^{'}X)^{-1}$$

resembles the covariance matrix

$$(X^{'}X)^{-1}$$

The simplest form of Poisson model using the square root link function is given as follows,

$$\mu = \eta_i^2 \; and \; Var(\mu) = \mu$$

The linear predictor is defined by

$$\eta(x) = \alpha + \beta(x - \mu)^2$$

The square root link

$$\eta(x) = \sqrt{\mu_i}$$

The information matrix under the Poisson model, is a given observation at x

$$I(\theta, x) = \upsilon(x) \left(\frac{\partial \eta(x)}{\partial \theta}\right)\left(\frac{\partial \eta(x)}{\partial \theta}\right)^{'}$$

where,

$$\frac{\partial \eta(x)}{\partial \theta} = \begin{bmatrix} 1 \\ (x_i - \mu)^2 \\ -2\beta(x_i - \mu) \end{bmatrix}$$

And

$$\upsilon(x) = \frac{1}{V(Y)} \left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 = \mu$$

Finding the point of maximum or minimum response, i.e. estimating the parameters, is a key application for this model. The optimal design is a set of points and weights that best optimize the selected criterion function; the criterion function is often connected to the precision of parameter estimations, such as the size of a confidence interval of parameter estimators' sum of the variance. The standardized information matrix is employed in the criterion function. The standardized information matrix for a certain design is the weighted total of the contributions from each of the n design points.

The D-Optimality is the most common criterion 'which seeks to maximize —X' X—, the Determinant of the information matrix (X' X) the design. This means that the optimal design matrix X* contains the n experiments which maximizes the determinant of (X' X). This connection between the design matrix and the determinant also explains the use of the "D" in the term D-optimal designs.

$$|X'X^*| = max(|X'X|)$$

Maximizing the determinant of the information matrix $(X'X)$ is equivalent to Minimizing the determinant of the dispersion matrix$(X'X)^{-1}$

Using such an idea, the D-efficiency of an arbitrary design, X, is naturally defined as

$$Eff\ D = \frac{|I(x,\beta)|}{|I(x,\beta)|}$$

$X^*$ is the true optimal design.

$$M(\theta,\xi) = \sum_i w_i v(x_i) \frac{\partial\eta(x)}{\partial\theta} \frac{\partial\eta(x)}{\partial\theta}'$$

$$\begin{bmatrix} 1 & (x_{ij}-\mu)^2 & -2\beta_j(x_i-\mu) \\ (x_{ij}-\mu)^2 & (x_{ij}-\mu)^4 & -2\beta(x_{ij}-\mu)^3 \\ -2\beta_j(x_i-\mu) & -2\beta(x_{ij}-\mu)^3 & 4\beta_j^2(x_{ij}-\mu)^2 \end{bmatrix}$$

There are several methods to the practice of determining the optimal design. These include algorithms, analytical, numerical and graphical methods. The method selected to derive D - optimal designs for the Gamma model is described below.
The derivation of D-optimal designs is been illustrated for three parameter sets: $\theta_A = (3, 1, 0)'$ ; $\theta_B = (3, -1, 0)'$ and $\theta_A C = (1, -3, 0)'$

Step 1: To find the symmetric design consisting of p=3 points, equal design weights is assumed.

$$\xi_3 = \begin{Bmatrix} x & 0 & -x \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{Bmatrix}$$

**The Suggested 3-point designs**

| True Parameters | Designs |
|---|---|
| $\theta_A = (3, 1, 0)'$ | $\xi_{3A} = \begin{Bmatrix} 1.79997 & 0 & -1.79997 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{Bmatrix}$ |
| $\theta_B = (3, -1, 0)'$ | $\xi_{3B} = \begin{Bmatrix} 0.9997 & 0 & -0.9997 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{Bmatrix}$ |
| $\theta_C = (1, -3, 0)'$ | $\xi_{3C} = \begin{Bmatrix} 0.6526 & 0 & -0.6526 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{Bmatrix}$ |

The optimal design is determined by the values of $\beta$ and therefore can be found if they are known. The linear predictor is

$$\eta(x) = \alpha + \beta(x-\mu)^2$$

Table 1. **Support points of the locally D-optimal design for the Poisson model for various values of $\mu_i$**

| $\mu_i$ | $\beta$ = 0.1 | $\beta$ = 0.3 | $\beta$ = 0.5 | $\beta$ = 0.7 | $\beta$ = 0.9 |
|---|---|---|---|---|---|
| 0.0050138 | 0.04193 | 0.12581 | 0.20968 | 0.29355 | 0.37743 |
| 0.0056112 | 0.04185 | 0.12557 | 0.20929 | 0.29301 | 0.37673 |
| 0.0168920 | 0.04041 | 0.12123 | 0.20206 | 0.28288 | 0.36371 |
| 0.0255272 | 0.03932 | 0.11796 | 0.19661 | 0.27525 | 0.35389 |
| 0.0266473 | 0.03918 | 0.11754 | 0.19590 | 0.27427 | 0.35263 |
| 0.0270363 | 0.03913 | 0.11739 | 0.19566 | 0.27393 | 0.35219 |
| 0.0387969 | 0.03767 | 0.11302 | 0.18837 | 0.26372 | 0.33907 |
| 0.0487017 | 0.03646 | 0.10940 | 0.18234 | 0.25528 | 0.32822 |
| 0.0625895 | 0.03481 | 0.10443 | 0.17405 | 0.24367 | 0.31330 |
| 0.0957185 | 0.03101 | 0.09303 | 0.15505 | 0.21708 | 0.27910 |
| 0.1049692 | 0.02999 | 0.08997 | 0.14995 | 0.20993 | 0.26990 |
| 0.1170383 | 0.02868 | 0.08604 | 0.14341 | 0.20077 | 0.25814 |
| 1.8832063 | 0.15143 | 0.45431 | 0.75719 | 1.06007 | 1.36295 |
| 1.8966063 | 0.15475 | 0.46426 | 0.77377 | 1.08328 | 1.39279 |
| 1.9084538 | 0.15771 | 0.47315 | 0.78858 | 1.10401 | 1.41945 |
| 1.9275791 | 0.16255 | 0.48767 | 0.81278 | 1.1379 | 1.46301 |
| 1.9406797 | 0.16591 | 0.49774 | 0.82957 | 1.16140 | 1.49323 |
| 1.9419884 | 0.16625 | 0.49875 | 0.83126 | 1.16376 | 1.49627 |
| 1.9488737 | 0.16803 | 0.50409 | 0.84016 | 1.17622 | 1.51229 |
| 1.9647796 | 0.17218 | 0.51654 | 0.86090 | 1.20527 | 1.54963 |

we plot the optimal design for varying $\beta$ . For a fixed $\alpha$, the curve represents linear predictors for various mean values as a function of x. The slope $\beta$ is changed to generate different linear predictors; we find that the optimal design decreases at first as the mean values increase, then reaches a zero at certain points, and then increases as the mean values increase. The optimal design has an equal allocation of [-1, 1].

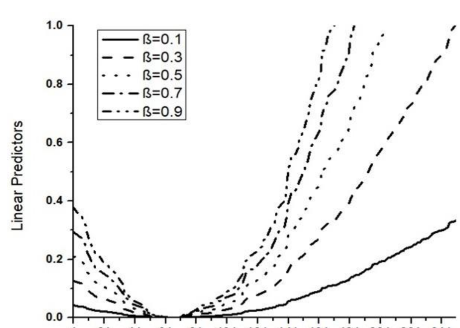


Figure 1

## 3. Conclusion

The solution to these formulas may be obtained numerically using R-software. The analytic contraction of D-optimal of Poisson model with square root link function was constructed to produce locally D-optimal design. As a result, the locally D-optimal design is a function of mean values and solely depends on the parameters. We find that in the best design with square root link function, as the mean values grow, the predicted value decreases steadily until the curve reaches zero, at which time the connection between function of mean values and predicted values increases. As it converges to one for various functional mean values, the linear predictor achieves maximum optimum values for each of the values.

## References


1. Biedermann, S., Yang. M: Abstract Wiener spaces, in: ***Designs for selected nonlinear models (2015). In Handbook of Design and Analysis of Experiments*** , (pp. 535-568). Chapman and Hall/CRC.
2. Chang,F.C, : D-optimal designs for weighted polynomial regression a functional-algebraic approach , ***Statistica Sinica 15(1),*** (2005a) 153-163.
3. Chang,F.C, : D-optimal designs for weighted polynomial regression a functional-algebraic approach , ***Annals of the Institute of Statistical Mathematics, 57(4), ,*** (2005a) 833-844.
4. Dette,H., Melas, V.B., and Pepelyshev, A. : Local c- and E-optimal designs for exponential regression," , ***Ann. Inst. Statist. Math.,58,*** 407-426.
5. Hang, Y , : Bayesian D-optimal design for generalized linear models , ***(Doctoral dissertation, Virginia Tech) ,*** (2006).
6. Russell, K. G., Woods, D. C., Lewis, S. M. and Eccleston, J. A.:D-optimal designs for Poisson regression models. ***Statistica Sinica 19,*** , (2009) 721 – 730.
7. Yang, M. and Stufken, J. .:On locally optimal designs for generalized linear models with group effects. ***Statistica Sinica 22,*** , 1765-1786



S.Joshua David: Department of Statistics, University of madras, Madras christian college chennai, Tamilnadu 600059,INDIA

*Email address*: 6joshuaadavid@gmail.com

Swetha N: Department of Statistics, University of madras, Madras christian college chennai, Tamilnadu 600059,INDIA

*Email address*: swethanagalingam29@gmail.com