EFFICIENCIES OF D- AND A-OPTIMAL DESIGNS FOR POISSON REGRESSION MODELS WITH TWO AND THREE VARIABLES

Supported by tetfund

E. I. Olamide*, F. B. Adebola and O. A. Fasoranbaku
Department of Statistics, Federal University of Technology, Akure – Ondo State, Nigeria
*Corresponding author: eiolamide@futa.edu.ng

Received: March 27, 2020 **Accepted:** August 19, 2020

Abstract: Efficiency plays an important role in the assessment of a design. An efficient design utilizes the design structure and information effectively. This study comparatively examines the design efficiencies of D- and A-optimal designs for Poisson regression models with two and three variables. The optimal designs were constructed for the considered models and their design efficiencies evaluated. The two-variable Poisson regression model was observed to yield greater efficiencies than the Poisson regression model with three variables. The A-optimality criterion produced greater design efficiencies both for the two- and three-variable Poisson regression models than the D-optimality criterion. The study recommends that the two-variable Poisson regression model should be broadly preferred to the three-variable Poisson regression model especially for A-optimality.

Keywords: A-optimality, D-optimality, design, information matrix, Poisson regression model

Introduction

Efficiency is a crucial phenomenon in optimal designs of experiments. A highly efficient design enhances good classification of optimal experimental designs through available data. Optimal designs aid exact estimation of model parameters by minimizing the variance of estimators; this also improves the design power.

Smith (1918) pioneered researches in the field of optimal experimental design through the formulation of a study on optimal experimental designs using mathematical approach. Smith (1918) formulated an innovative mathematical study on design of experiments and this was later extended by Kiefer and Wolfowitz (1959) when the formation of a “theoretical” framework aimed at obtaining optimal designs based on criteria through the expression of design as a measure of probability that represents treatment allocations to any specific region in the design space was considered. The methodology formed, and the discussions on D- and E-optimality criteria concerning a regression model in linear terms set the foundation for other design criteria such as the A-, F-, G- and Q-optimality criteria.

Before data collection is considered, optimal designs can be strategized or designs can be premeditated such that collection of data is done consecutively. In either procedure, an experiment that is cautiously designed, can give precise statistical inference at very low cost and with great efficiency. Berger and Wong (2009) demonstrated that the number of observations when optimal designs are constructed is being reduced by 20 to 40% as compared to classical designs.

Determinant of Fisher information matrix was used to study designs for D-optimality comparatively (Wald, 1943). Silvey (1980) presented explicit descriptions of the most frequently used design criterion in optimization problems.

Aguiar *et al*. (1995) noted that several classical symmetrical designs exhibit required features, one of which is the D-optimality. Application of the notion of D-optimality can be extended to the selection of design when the classical symmetrical designs cannot be used, for instance, when there is irregularity in the shape of experimental region, when dealing with very large experimental runs for a classical design the number of experiments chosen by a classical design is too large or when one considers application of models that differ from that of the usual first or second order.

Bingham and Chipman (1996) introduced a design criterion that searches for the maximization of discriminating link between models. The Hellinger distance between predictive distributions in competing models, inspired by Meyer *et al*. (1996) is the basis on which the criterion is found. The bound of the criterion that critically increases improvement in interpretability was obtained. The set of all possible models to be compared was noted to be massive, observing that all models were not possible alike. A Bayesian approach was employed in addressing the challenge. The method contains prior distributions on the space of models, signifying inclination for models that are attractive instinctively. Typical examples of such models are the ones whose effects are few, models with more of low order than high order effects and models with inheritance structure between active main effects and interactions. Procedures of assessment of the criterion and search of optimal designs were presented. The significant effects of regular and non-regular designs, robust designs, as well as scenarios with limited previous knowledge were considered to illustrate the efficiency of the design criterion through some examples.

Boer and Hendrix (2000) showed the formation of an inspiring applicable area of global optimization through optimal experimental designs. The structures of challenging global optimisation problems and the corresponding procedures of determining such structures through optimal designs were considered. Discussion was majorly on three types of designs; these include the exact designs, replication-free designs and discrete designs. Generation of optimal designs for the three notions was observed to comprise different optimisation problems.

Dror and Steinberg (2005) proposed a quick and easy technique that aids the construction of approximate locally D-optimal designs with great efficiencies for multivariate models containing binary responses. D-optimal designs were obtained through theapproach for similar problem having a normally distributed response containing the same linear predictor, with a supposition of homogeneity in variance.The required change for the transformation of the standard design into an efficient one using multivariate logit or probit model is shifting a design point that has too low or very high probability to the nearest possible point of moderate probability.

Debusho and Haines (2010) considered a simple linear regression model with random coefficient for generating D-optimal designs. The factors were extracted from a group of time points of equal spacing and infrequence. The dependence of the designs on variance-component values was examined. Linear transformation of the time points does not necessarily guarantee the mapping onto one another of the designs when population with both fixed effects and variance components were considered. Numerical illustration was presented through an example to validate the finding.

Schorning *et al*. (2017) considered an actively controlled clinical dose discovery experiment for the derivation of optimal designs in estimating the efficacy and toxicity when the bivariate continuous outcomes are modeled either by second order polynomial, Michaelis-Menten model, Emax model, or a mixture of any two. The sufficient conditions for the addition of boundary points of the design space in the optimal design were provided and higher bounds of the number of diverse dose levels necessary for the optimal design were obtained. The independence of the minimally supported D-optimal designs on the correlation between bivariate outcomes was analytically described. Demonstration of the proposed methods through numerical examples was illustrated and the advantage of the D-optimal designs was also demonstrated using experiment that was lately examined in the literature.

This research evaluates the efficiencies of the D- and A-optimal criteria for Poisson regression models with two and three variables.

Materials and Methods

Poisson regression model

Poisson regression model can be broadly written as follows:

$$y_{ij} \sim Poisson(\tau_i)$$

The mean response, τ_i, can as well be expressed as follows:

$$\tau_i = exp(X_i'\beta)$$

Where: y_{ij} are the response variables, τ_i is the expectation of the response variable at the i^{th} design point X_i' is the design matrix containing factors X_i (i = 1, 2, …), β is a vector of parameters

Construction of D-optimal designs for Poisson regression models with two and three variables

The two- and three-variable Poisson regression models can be represented by equations (2.1) and (2.2), respectively;

$$\tau_i = exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}) \quad (2.1)$$

and

$$\tau_i = exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i}) \quad (2.2)$$

The D-optimum design searches for the maximization of the determinant of the Fisher information matrix or equivalent lysearches for the minimization of the determinant of the inverse of the Fisher information matrix.

Mathematically, if the dimension of β is $p \times 1$, the Fisher information matrix, $I(X,\beta)$, is a $p \times p$ matrix as presented in equation (2.3)

$$I(X,\beta) = -E\left[\frac{\partial^2 \log(L(X, \beta))}{\partial\beta\partial\beta'}\right] \quad (2.3)$$

Where: $(L(X, \beta)$ is the likelihood function of the data, X, is the design matrix.

The commonly used D-optimal criterion is defined by equation (2.4)

$$\max_{X \in \mathcal{D}} det\left[\frac{I(X,\beta)}{n}\right], \quad (2.4)$$

Where: n is the total sample size, $\boldsymbol{\mathcal{D}}$ is the set of all possible designs.

Since in most cases, n is fixed, the D-optimal design is obtained by maximizing the determinant of the Fisher information matrix.

The information matrix is expressly defined in terms of a design measure as $M(\xi; \beta)$.

D-optimal = min[| (X'X) $^{-1}$|] or max[(X'X)].

For the two-variable Poisson regression model in equation (2.1), the transformation via the natural logarithm and first derivative are presented in equations (2.5) and (2.6) respectively.

$$ln\, \tau_i = \eta_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} \quad (2.5)$$

Here,

$$f'(x_i) = (1, \quad x_1, \quad x_2) \quad (2.6)$$

The elements of the information matrix are thus obtained as equation (2.7):

$$M = X'X = \begin{bmatrix} 1 & x_1 & x_2 \\ x_1 & {x_1}^2 & x_1 x_2 \\ x_2 & x_1 x_2 & {x_2}^2 \end{bmatrix}, \quad (2.7)$$

The Fisher information matrix can be expressed in a compact form as equation (2.8):

$$M(\xi; \beta_0, \beta_1, \beta_2) = \sum w_i \tau_i f(x_i) f'(x_i) \quad (2.8)$$

and more compactly, as equation (2.9)

$$M(\xi; \beta_0, \beta_1, \beta_2) = X'WX \quad (2.9)$$

Where: w_i denote the weights of the support points, $\tau_i = \exp(\eta_i)$, is the mean response of the i^{th} design point, ξ is the design measure, $W = diag\{w_i \tau_i\}$, $X = [f(x_1), f(x_2)]$

Explicitly, the Fisher information matrix can therefore be expressed as equation (2.10):

$$M(\xi; \beta_0, \beta_1, \beta_2) = \begin{bmatrix} \sum w_i\tau_i & \sum w_i\tau_i x_{1i} & \sum w_i\tau_i x_{2i} \\ \sum w_i\tau_i x_{1i} & \sum w_i\tau_i {x_{1i}}^2 & \sum w_i\tau_i x_{1i}x_{2i} \\ \sum w_i\tau_i x_{2i} & \sum w_i\tau_i x_{1i}x_{2i} & \sum w_i\tau_i {x_{2i}}^2 \end{bmatrix} \quad (2.10)$$

The D-optimal design, ξ^*, for the two-variable Poisson regression model is that function that satisfies equation (2.11)

$$|M(\xi^*; \beta_0, \beta_1, \beta_2)| = \max_{\xi \in \Xi} |M(\xi; \beta_0, \beta_1, \beta_2)| \quad (2.11)$$

To obtain the efficiency with regards to D-optimal design for the model, the efficiency of design, ξ, is measured relative to design, ξ^*. The D-efficiency of a random design, ξ, is thus defined as equation (2.12a):

$$D_{eff} = \left(\frac{|M(\xi; \beta_0, \beta_1, \beta_2)|}{|M(\xi^*; \beta_0, \beta_1, \beta_2)|}\right)^{1/p} \quad (2.12a)$$

Alternatively, the D-efficiency can be computed by equation (2.12b)

$$D - efficiency = 100 * [(|X'X|1/p)/N] \quad (2.12b)$$

Where: *p* represents the number of parameter and *N*is the number of experimental runs.

For the three-variable Poisson regression model in equation (2.2), the information matrix is obtained as equation (2.13);

$$M(\xi,\beta_0,\beta_1,\beta_2,\beta_3) = \begin{bmatrix} \sum_{i=1}^{4} w_i\tau_i & \sum_{i=1}^{4} w_i\tau_i\, x_{1i} & \sum_{i=1}^{4} w_i\tau_i\, x_{2i} & \sum_{i=1}^{4} w_i\tau_i\, x_{3i} \\ \sum_{i=1}^{4} w_i\tau_i\, x_{1i} & \sum_{i=1}^{4} w_i\tau_i\, {x_{1i}}^2 & \sum_{i=1}^{4} w_i\tau_i\, x_{1i}x_{2i} & \sum_{i=1}^{4} w_i\tau_i\, x_{1i}x_{3i} \\ \sum_{i=1}^{4} w_i\tau_i\, x_{2i} & \sum_{i=1}^{4} w_i\tau_i\, x_{1i}x_{2i} & \sum_{i=1}^{4} w_i\tau_i\, {x_{2i}}^2 & \sum_{i=1}^{4} w_i\tau_i\, x_{2i}x_{3i} \\ \sum_{i=1}^{4} w_i\tau_i\, x_{3i} & \sum_{i=1}^{4} w_i\tau_i\, x_{1i}x_{3i} & \sum_{i=1}^{4} w_i\tau_i\, x_{2i}x_{3i} & \sum_{i=1}^{4} w_i\tau_i\, {x_{3i}}^2 \end{bmatrix} \quad (2.13)$$

The D-optimal design for the model in equation (2.2) is that function that satisfies equation (2.14)

$$|M(\xi^*;\ \beta_0,\beta_1,\beta_2,\beta_3)| = \max_{\xi\in\Xi}|M(\xi;\ \beta_0,\beta_1,\beta_2,\beta_3)| \quad (2.14)$$

In the same vein, the D-efficiency for the model is obtained by equation (2.15):

$$D_{eff} = \left(\frac{|M(\xi;\beta_0,\beta_1,\beta_2,\beta_3)|}{|M(\xi^*;\ \beta_0,\beta_1,\beta_2,\beta_3)|}\right)^{1/p} \quad (2.15)$$

Construction of A-optimal designs for Poisson regression models with two and three variables

TheA-optimality design criterion seeks the minimization of the trace relating to the inverse of the information matrix. In other words, the sum of the variances of the parameter estimates is minimized, equivalently minimizing the average variance.

For the model in equation (2.1),with information matrix presented in equation (2.10), the A-optimal design can be computed via equation (2.16)

$$A-optimal = \min[tr(M(\xi;\ \beta_0,\beta_1,\beta_2))] \quad (2.16)$$

In terms of eigenvalues, suppose the information matrix in equation (2.10) has eigenvalues, λ_i, then, the expression for A-optimality is as presented in equation (2.17)

$$A-optimal = \min\sum_{i=1}^{p}\frac{1}{\lambda_i} \quad (2.17)$$

Where: λ_i are the eigen values of the information matrix, and p is the number of parameters.

The efficiency of A-optimal design can be calculated by equation (2.18)

$$A-\text{efficiency} = 100 * [p/\text{trace}(N * (\mathbf{X'X})^{-1})] \quad (2.18)$$

Results and Discussion

D-optimal designs for Poisson regression models with two and three variables

The results of D-optimal designs for Poisson regression models with two and three predictor variables in linear terms are presented in this section.

Considering the two-variable Poisson regression model in equation (2.1), the assumption is that $x_{1i}, x_{2i} \in [0,1]$ $(i = 1,\cdots,p+1)$, and $\beta = (1,\ 2,\ -2)^T$.

At 3-design points, the generated D-optimal design is presented in equation (3.1a)

$$\xi_D^* = \begin{Bmatrix} (0, & 0)(1, & 1)(1, & 0) \\ & \frac{1}{3}\frac{1}{3}\frac{1}{3} & \end{Bmatrix} \quad (3.1a)$$

Equation (3.1a) shows the D-optimal designsconstructed for the multiple linear Poisson regression model pertaining to two predictor variables. It thus means that if there are 100 experimental runs, 33.33% of the total units should be allocated to design points $x_1 = 0$and $x_2 = 0$, $x_1 = 1$ and $x_2 = 1$, and $x_1 = 1$ and $x_2 = 0$, respectively.

At 4-design points, the D-optimal design constructed for the multiple linear Poisson regression model in two variables using the same constrained design space is presented in equation (3.1b);

$$\xi_D^* = \begin{Bmatrix} (0,\ 0)(0,\ 1)(1,\ 0)(1,\ 1) \\ \frac{1}{4}\frac{1}{4}\frac{1}{4}\frac{1}{4} \end{Bmatrix} \quad (3.1b)$$

Considering equation (3.1b), the design for the multiple linear Poisson regression model in two variableis also D-optimal at a collection of 4 design points $x_1 = 0$ and $x_2 = 0$, $x_1 = 0$ and $x_2 = 1$, $x_1 = 1$ and $x_2 = 0$ and $x_1 = 1$ and $x_2 = 1$. Each group of design points has design weight of 0.25.

Consideringthe Poisson regression model with three design variables in equation (2.2), the generated D-optimal designs is presented in equation (3.2);

$$\xi_D^* = \begin{Bmatrix} (1,\ 0,\ 1)(0,\ 0,\ 0)(1,\ 1,\ 1)(0,\ 0,\ 1) \\ \frac{1}{4}\frac{1}{4}\frac{1}{4}\frac{1}{4} \end{Bmatrix} \quad (3.2)$$

Using $\beta = (4,\ 1,\ 3,\ 2)^T$ as the best vector of parameter guess, equation (3.2) shows that the designsconstructed for the multiple linear Poisson regression model with three predictor variables is D-optimal at 4-design points $x_1 = 1$, $x_2 = 0$, $x_3 = 1$; $x_1 = 0$, $x_2 = 0$, $x_3 = 0$; $x_1 = 1$, $x_2 = 1$, $x_3 = 1$; and $x_1 = 0$, $x_2 = 0$, $x_3 = 1$. This means that, in an experimental layout, 25% of the total experimental runs are allocated to each optimal design region.

A-optimal designs for Poisson regression models with two and three variables

The results of A-optimal design pertaining to a Poisson regression model with two predictor variables is hereby presented and discussed. For Poisson regression model involving two predictor variables, the constructed A-optimal design is presented in equation (3.3a)

$$\xi_A^* = \begin{Bmatrix} (0, & 0)(0, & 1)(1, & 0)(1, & 1) \\ & \frac{1}{4}\frac{1}{4}\frac{1}{4}\frac{1}{4} & \end{Bmatrix} \quad (3.3a)$$

The design is observed to be A-optimal at 4-design points. After 1000 iterations, the A-optimal criterion value is 100 and the optimal design points are $x_1 = 0$, $x_2 = 0$; $x_1 = 0$, $x_2 = 1$; $x_1 = 1$, $x_2 = 0$ and $x_1 = 1$, $x_2 = 1$. The constructed A-optimal design weight at each optimal design pointis $w_1 = w_2 = w_3 = w_4 = 0.25$.

Considering the model in equation (2.2), the constructed A-optimal design is presented in equation (3.3b)

$$\xi_A^* = \begin{Bmatrix} (0,0,1)(0,1,0)(0,1,1)(1,0,0)(1,1,0)(1,1,1) \\ \frac{1}{6}\frac{1}{6}\frac{1}{6}\frac{1}{6}\frac{1}{6}\frac{1}{6} \end{Bmatrix} \quad (3.3b)$$

Considering the three-variable Poisson regression model, the design in equation (3.3b) is A-optimal at 6-design points with support points $x_1 = 0,\ x_2 = 0,\ x_3 = 1$; $x_1 = 0,\ x_2 = 1,\ x_3 = 0 x_1 = 0,\ x_2 = 1,\ x_3 = 1$; $x_1 = 1,\ x_2 = 0,\ x_3 = 0$; $x_1 = 1,\ x_2 = 1,\ x_3 = 0$; and $x_1 = 1,\ x_2 = 1,\ x_3 = 1$. The constructed A-optimal design weight at each optimal design pointis $w_1 = w_2 = w_3 = w_4 = w_5 = w_6 = 0.1667$.

Table 1: D-optimal design efficiencies

Model	Efficiency Lower Bound (ELB)
$\tau_i = exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i})$	75.0000%
$\tau_i = exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i})$	71.4286%

Table 2: A-optimal design efficiencies

Model	A-Efficiencies
$\tau_i = exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i})$	100%
$\tau_i = exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i})$	88%

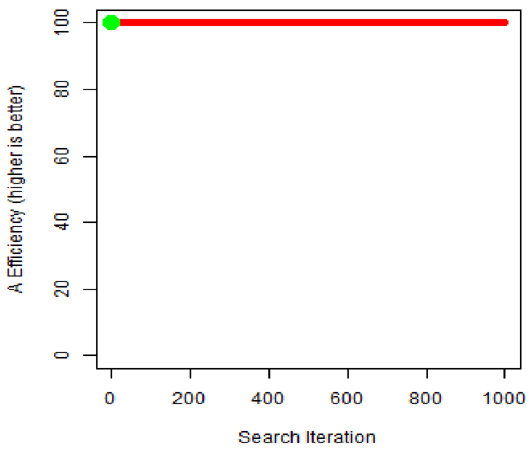

Fig. 1: Optimal efficiency search values of A-optimal design for Poisson regression model with two variables

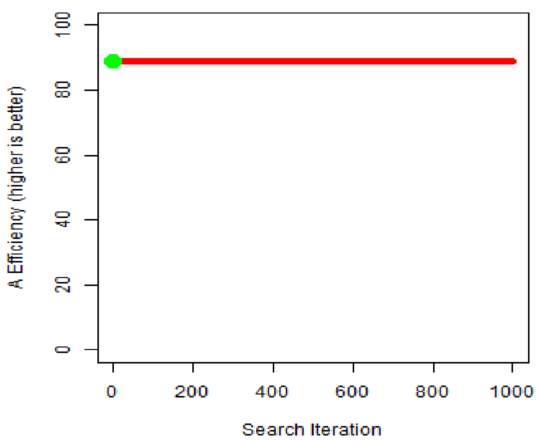

Fig. 2: Optimal efficiency search values of A-optimal design for Poisson regression model with three variables

Efficiencies of D- and A-optimal designs

Tables 1 and 2 show the efficiencies of D- and A-optimal designs for two- and three- variable Poisson regression models respectively.The modelsin equations (2.1) and (2.2) were found to be 75% D-efficient and 71.43% D-efficient, respectively. For A-efficiencies, the models in equations (2.1) and (2.2) were found to be 100% A-efficient and 88% A-efficient respectively. The results of the A-efficiencies can also be observed from Figs. 1 and 2, respectively.

Conclusion

This study examines the efficiencies of D- and A-optimal designs for Poisson regression models involving two and three explanatory variables. Optimal designs were constructed for the considered criteria and their efficiencies were evaluated. The optimality criteria considered were found to be more efficient for two-variable Poisson regression than three-variable Poisson regression. The two-variable Poisson regression model was observed to be perfectly A-efficient, yielding 100% design efficiency. Comparatively, this implies that the two-variable Poisson regression model should be broadly preferred to the three-variable Poisson regression model.

Conflict of Interest

Authors declare that there is no conflict of interest reported in this work.

References

Berger MPF & Wong W 2009. An Introduction to Optimal Designs for Social and Biomedical Research. *Wiley*.

Bingham DR & Chipman HA 1996. Optimal Designs for Model Selection.

Boer EPJ & Hendrix EMT 2000. Global optimization problems in optimal design of experiments in regression models. *Journal of Global Optimization*, 18(4): 385-398.

deAguiar PF, Bourguignon B, Khots MS, Massart DL & Phan-Than-Luu R 1995. D-optimal Designs. *Chemometrics and Intelligent Laboratory Systems*, 30(2): 199-210.

Debusho LK & Haines LM 2010. D-optimal population designs for the simple linear random coefficients regression model. *Annual Proceedings of the South African Stat. Assoc. Conf.*, Congress, 1: 52-59.

Dror HA & Steinberg DM 2005. Approximate local D-optimal Experimental Design for Binary Response, *Technical Report* RP-SOR-0501, Tel. Aviv University.

Kiefer J & Wolfowitz J 1959. Optimum designs in regression problems. *Annals of Mathematical Statistics*, 30: 271-294.

Smith K 1918. On the standard deviations of adjusted and interpolated values of an observed polynomial function and its constants and the guidance they give towards a proper choice of the distribution of observations. *Biometrika*, 12: 1-85.

Wald A 1943. On the efficient design of statistical investigations. *Ann. Math. Stat.*, 14(2): 134-140.

Schorning K, Dette H, Kettelhake K, Wong WK & Bretz F 2017. Optimal designs for active controlled dose-finding trials with efficacy-toxicity outcomes. *Biometrika*, 104(4): 1003-1010.

Silvey D 1980. Optimal Design. *Chapman and Hall*, London.