Journal of Asian Scientific Research, 2014, 4(3): 120-124

Journal of Asian Scientific Research

journal homepage: http://www.aessweb.com/journals/5003

D-OPTIMAL DESIGN FOR LOGISTIC REGRESSION MODEL WITH THREE INDEPENDENT VARIABLES

Habib Jafari

Faculty Member of Razi University, Department of Statistics, Kermanshah-Iran

Soliman Khazai

Faculty Member of Razi University, Department of Statistics, Kermanshah-Iran

Yazdan Khaki

M.Sc. in statistic, Razi University of Kermanshah-Iran

Tohid Jafari

Department of Technical and Engineering, Tabriz Branch, Islamic Azad University, Tabriz, Iran

ABSTRACT

Non-linear models are of particular importance owing to their various applications in different fields. In this paper, considering D-optimal criterion, appropriate models were introduced and, based on a logistic regression model with three independent variables and dependence of information matrix on passive parameters, a locally D-optimal design was obtained for several specific states. It is noteworthy that certain designs with different points were presented and their subject optimality was calculated based on space of the parameters.

Keywords: Locally D-optimal criterion, Locally D-optimal design, Information matrix, Logistic regression, Independent variables, Support points.

1. INTRODUCTION

Model estimating and fitting are two important topics in statistical sciences. The present paper was written considering the crucial role of statistical discussions in other sciences such as industries, medicine and so on. As mentioned earlier, model fitting is one of the important discussions in statistics; thus, in most cases, fitting and obtaining a relationship between dependent variable(s) and independent variable(s) are intended to be achieved. When the effect of an independent variable on a dependent variable is investigated, two points should be taken into consideration: a) finding independent variables which affect dependent variables, and b) finding values of dependent variables which have a significant role in model fitting. Therefore, the basic motivation for designing optimal experiments is to find an ideal design that could result in appropriate inferences about model parameters by conducting an experiment using this design.

In statistical discussions, models are generally divided into two groups: linear models and nonlinear models.

Non-linear models are frequently utilized in medicine and pharmacy. Although optimal designs have been used for linear models for years ago and many results have been extracted in this regard, few results and findings have been obtained from non-linear models due to their complex applications. Accordingly, in this paper, an optimal design was obtained from logistic regression model with three independent variables. Thus far, extensive studies have been conducted on finding a D-optimal design for linear and non-linear models, which include Fedorove and Hackle [1], Atkinson and Donove [2], Haines and Gaetan [3], Li and Majumdar [4] and Chao [5].

The remaining of this paper is organized into three sections. The first section includes the statistical model and its characteristics, optimal design is described in the third section. And the last section is discussed.

2. STATISTICAL MODEL AND ITS CHARACTERISTICS

In this section, the utility function which includes two parts, knows the regression function with three independents variables and error term, is written as follows:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i; \qquad i = 1,2,\cdots,n.$$

Since response variable y has Bernoulli distribution and link function of the model is considered logistic regression with three independent variables, therefore:

$$\pi_i = \frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\beta_3 x_{i3}}}{1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\beta_3 x_{i3}}}; \qquad i = 1,2,\cdots,n. \tag{1}$$

where π_i is success probability, β_0 ,β_1,β_2 , β_3 is unknown parameters and x_1, x_2 , $x_3$indicates independent variables which accept positive values. It is assumed that β_0 ,β_1,β_2 , $\beta_3$belongs to real numbers. Then, considering the studied issue, the need to generate D-optimality criterion which is a function in terms of information matrix of a design with m points and according to Relation (1), likelihood function for a single observation (x_{i1}, x_{i2}, x_{i3}) will be as follows:

$$L(\boldsymbol{\beta}; \boldsymbol{y}) = \pi_i^{y_i}(1-\pi_i)^{1-y_i} = \left(\frac{\pi_i}{1-\pi_i}\right)^{y_i}(1-\pi_i) \tag{2}$$

Considering the likelihood function in Relation (2) and mathematical expectation of second-order partial derivatives, this function is calculated in terms of the vector of parameters, Fisher's information matrix related to parameters $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)$and vector $\boldsymbol{x}_i = (x_{1i}, x_{2i}, x_{3i})$as follows:

$$\mathbf{M}(\boldsymbol{\beta}; \boldsymbol{x_i}) = \frac{e^u}{1+e^u}\begin{pmatrix} 1 & x_{1i} & x_{2i} & x_{3i} \\ x_{1i} & x_{1i}^2 & x_{1i}x_{2i} & x_{1i}x_{3i} \\ x_{2i} & x_{1i}x_{2i} & x_{2i}^2 & x_{2i}x_{3i} \\ x_{3i} & x_{1i}x_{3i} & x_{2i}x_{3i} & x_{3i}^2 \end{pmatrix}$$

where $u = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3}$.

3. OPTIMAL DESIGN

Considering the introduced model which consists of three independent variables x_1, x_2 , x_3, design (3) is defined in order to estimate the model parameters:

$$\xi = \begin{pmatrix} \begin{pmatrix} x_{11} \\ x_{21} \\ x_{31} \end{pmatrix} & \begin{pmatrix} x_{12} \\ x_{22} \\ x_{32} \end{pmatrix} & \cdots & \begin{pmatrix} x_{1m} \\ x_{2m} \\ x_{3m} \end{pmatrix} \\ w_1 & w_2 & & w_m \end{pmatrix} \in \Xi \qquad (3)$$

where $\Xi = \{\xi | \xi \in \Xi\}$ and $x_{ij};\ i = 1,2,3$ and $j = 1,2,\dots,m$ demonstrate points related to vectors **X** and $\sum_{i=1}^{m} w_i = 1,\ \ 0 \le w_i \le 1$, respectively. Considering that information matrix of these models (non-linear) depends on the vector of parameters, the number of points of the defined design (3) will be as follows:

$$p \le m \le \frac{p(p+1)}{2}$$

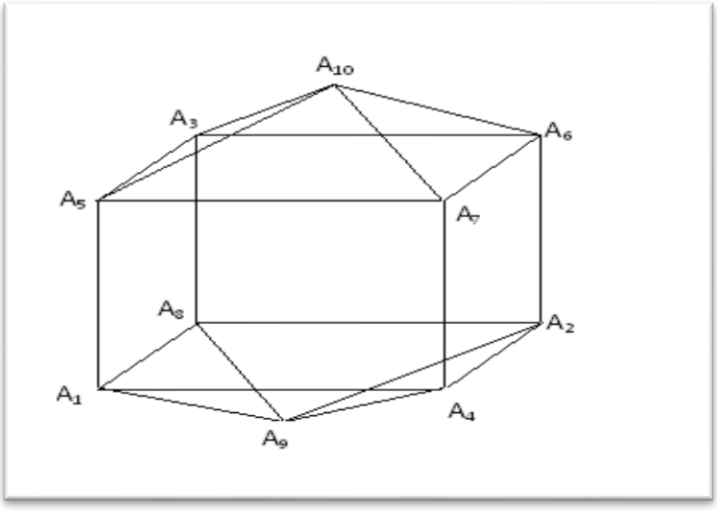
Therefore, considering the aforesaid inequality and the number of model parameters $p = 3,4,\dots,10$ point designs could be defined in this state as in Fig.1; each of them can be optimized in a certain area of parameter space based on reliance on passive parameters. Accordingly, a 10-point design was defined in this article, in which, considering the complexity of the design, constant supporting points cannot be obtained (at least using available software) and thus variable change was used. For example, a 10-point design, ξ, in a specific case, can be defined as follows:

$$\xi = \begin{pmatrix} \begin{pmatrix} u-\beta_0 \\ 0 \\ 0 \end{pmatrix} & \begin{pmatrix} 0 \\ u-\beta_0 \\ 0 \end{pmatrix} & \begin{pmatrix} 0 \\ 0 \\ \frac{3(u-\beta_0)}{2} \end{pmatrix} & \begin{pmatrix} u-\beta_0 \\ u-\beta_0 \\ \frac{(u-\beta_0)}{2} \end{pmatrix} & \begin{pmatrix} u-\beta_0 \\ 0 \\ \frac{3(u-\beta_0)}{2} \end{pmatrix} & \begin{pmatrix} 0 \\ u-\beta_0 \\ \frac{3(u-\beta_0)}{2} \end{pmatrix} & \begin{pmatrix} u-\beta_0 \\ u-\beta_0 \\ \frac{3(u-\beta_0)}{2} \end{pmatrix} & \begin{pmatrix} 0 \\ 0 \\ \frac{(u-\beta_0)}{2} \end{pmatrix} & \begin{pmatrix} \frac{(u-\beta_0)}{2} \\ \frac{(u-\beta_0)}{2} \\ 0 \end{pmatrix} & \begin{pmatrix} \frac{(u-\beta_0)}{2} \\ \frac{(u-\beta_0)}{2} \\ 2(u-\beta_0) \end{pmatrix} \\ w_1 & w_2 & w_3 & w_4 & w_5 & w_6 & w_7 & w_8 & w_9 & w_{10} \end{pmatrix} \in \Xi \quad (4)$$

where $0 < u \le \beta_0$.

In Fig.1, a geometric graph is drawn that demonstrates state of the defined points in the design (4).

Figure- 1. Graph related to design (4) and the A1 to the design note the points of the design (4).



As mentioned in the abstract, to obtain an optimal design, D-optimality criterion which is a linear function based on the matrix model is utilized as follows:

$$\mathrm{D}-\mathrm{Criterion}=\psi(\xi)=-\ln\det\left(\mathbf{M}(\boldsymbol{\beta};\xi)\right)$$

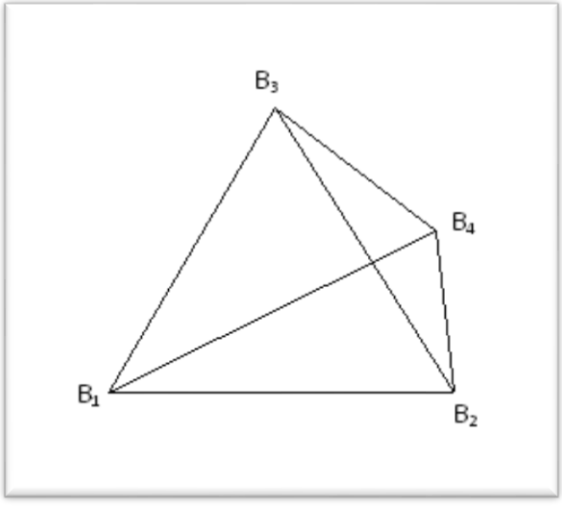
where ξ is the desired design and $\mathbf{M}(\xi)$ is Fisher's information matrix relating to design ξ, which is parameter-dependent on linear models and is calculated as shown below:

$$\boldsymbol{M}(\boldsymbol{\beta};\xi)=\sum_{i=1}^{10}w_i\boldsymbol{M}(\boldsymbol{\beta};x_i)$$

where $\boldsymbol{M}(\boldsymbol{\beta};x_i)$ is the information matrix related to m^{th} vector from dependent variables. Considering that information matrix of the considered model is parameter-dependent, based on the number of definable points of the design, ($p\le m\le\frac{p(p+1)}{2}$), it is possible to find designs with different points considering the space of parameters so that a specific type (of the number of points) of this design can be optimized in each subspace of parameters (partition). In other words, a design with only 4 points can be optimized in one partition, a 5-point design can be optimized in another partition and so on. Therefore, a 10-point design can be defined for a specific state (4), in which, for $\beta_0=-20,\dots,+5$, 10-point design is converted into a 4-point design with equal weights as shown below:

$$\xi_1=\begin{pmatrix}\begin{pmatrix}u-\beta_0\\0\\0\\\frac{1}{4}\end{pmatrix}&\begin{pmatrix}0\\u-\beta_0\\0\\\frac{1}{4}\end{pmatrix}&\begin{pmatrix}0\\0\\\frac{3(u-\beta_0)}{2}\\\frac{1}{4}\end{pmatrix}&\begin{pmatrix}u-\beta_0\\u-\beta_0\\\frac{3(u-\beta_0)}{2}\\\frac{1}{4}\end{pmatrix}\end{pmatrix} \tag{5}$$

Figure-2. Graph related to the design (5).



Determinant of information matrix of design (5) is calculated as follows:

$$det\left(\boldsymbol{M}(\boldsymbol{\beta};\xi)\right)=\frac{9e^{4u}(u-\beta_0)^6}{256(1+e^u)^8}$$

Considering values of $\beta_0$and D-optimality criterion, certain values of u are found which minimize the mentioned criterion. Table 1 demonstrates values of u and D for 4-point design (5).

Table-1. The values of $\boldsymbol{u}^*$ And D for differentiation of $\boldsymbol{\beta_0}$ in design (5).

β_0	-20	-3.5	-3	-2.5	-2	-1.5	-1	0	0.5	1
u^*	0.149	0.74	0.83	0.93	1.68	1.23	1.43	1.98	2.32	2.7
D	-9.1	0.76	1.5	2.7	3.62	4.65	5.82	8.6	10.14	11.84

For instance, for $\beta_0 = -3$, D-optimal design $\xi_1^*$is written as follows:

$$\xi_1^* = \begin{pmatrix} \begin{pmatrix} 3.83 \\ 0 \\ 0 \end{pmatrix} & \begin{pmatrix} 0 \\ 3.83 \\ 0 \end{pmatrix} & \begin{pmatrix} 0 \\ 0 \\ 5.745 \end{pmatrix} & \begin{pmatrix} 3.83 \\ 3.83 \\ 5.745 \end{pmatrix} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix} \quad (6)$$

In a specific state, for $\beta_0 = -3$, design (6) is locally D-optimal; since the number of points of the design is equal to the number of model parameters, thus, equal weights are assigned to the obtained design points.

4. DISCUSSION AND CONCLUSION

In this paper, considering that locally D-optimal model was calculated for logistic regression model with three independent variables in a specific state, the idea of building locally optimal designs for logistic regression models without random effects was presented. Among the results of this discussion is that: if β_0 changes between -20 and +5, the appropriate design with 4 points would be optimal and other designs such as 5-, 6- and 10-point designs could be optimized in other partitions of the parameter.

5. ACKNOWLEDGEMENT

This work is partially supported by the Cooperative 4407 knowledge Base New Ideas.

REFERENCES

[1] V. V. Fedorove and P. Hackle, *Model-oriented design of experiments, Lecture note in statistics*: Springer-Verlag, 1972.

[2] A. Atkinson and A. Donove, *Optimum experimental designs, with SAS*. Oxford: Oxford University Press, 2007.

[3] L. Haines and K. M. Gaetan, *D-optimal designs for logistic regression in two variables*. South Africa: Department of statistical sciences university of Cape Town South Africa. School of Statistics and Actuarial Science, University of KwaZulu-Natal, Pietermaritzburg 3200, 2007.

[4] G. Li and D. Majumdar, *D-optimal designs for logistic models with three and four parameters*. USA: Glaxo Smith Kline, Collegeville, PA 19426, USA. University of Illionois at Chicago, IL 60607, 2007.

[5] Z. Chao, "Construction of optimal designs in polynomial regression models," M.Sc Thesis, 2012.