D-optimal designs for logistic regression in two variables

Linda M. Haines[1], M. Gaëtan Kabera[2], P. Ndlovu[2] and T. E. O'Brien[3]

[1] Department of Statistical Sciences, University of Cape Town, Rondebosch 7700, South Africa. lhaines@stats.uct.ac.za
[2] School of Statistics and Actuarial Science, University of KwaZulu-Natal, Pietermaritzburg 3200, South Africa. ndlovup@ukzn.ac.za
[3] Department of Mathematics and Statistics, Loyola University Chicago, Chicago, Illinois 60626, U.S.A. teobrien@gmail.com

Summary. In this paper locally D-optimal designs for the logistic regression model with two explanatory variables, both constrained to be greater than or equal to zero, and no interaction term are considered. The setting relates to dose-response experiments with doses, and not log doses, of two drugs. It is shown that there are two patterns of D-optimal design, one based on 3 and the other on 4 points of support, and that these depend on whether or not the intercept parameter β_0 is greater than or equal to a cut-off value of -1.5434. The global optimality of the designs over a range of β_0 values is demonstrated numerically and proved algebraically for the special case of the cut-off value of β_0.

Key words: D-optimality. Logistic regression in two variables.

1 Introduction

Logistic regression models with two or more explanatory variables are widely used in practice, as for example in dose-response experiments involving two or more drugs. There has however been only sporadic interest in optimal designs for such models, with the papers of [Sitter and Torsney (1995)], [Atkinson and Haines (1996)], [Jia and Myers (2001)], [Torsney and Gunduz (2001)] and [Atkinson (2006)] and the thesis of [Kupchak (2000)] providing valuable insights into the underlying problems. In the present study a simple setting, that of the logistic regression model in two explanatory variables with no interaction term, is considered. The variables are taken to be doses, and not log doses, of two drugs and are thus constrained to be greater than or equal to zero. The aim of the study is to construct locally D-optimal designs, and in so doing to identify patterns in the designs that may depend on the values of the parameters in the model, and in addition to demonstrate the global optimality of these designs both numerically and algebraically.

2 Preliminaries

Consider the logistic dose-response model defined by

$$\text{logit}(p) = \beta_0 + \beta_1 d_1 + \beta_2 d_2$$

where p is the probability of success, β_0, β_1 and β_2 are unknown parameters and d_1 and d_2 are doses, not log doses, of two drugs such that $d_1 \geq 0$ and $d_2 \geq 0$. Responses are assumed to increase with dose for both drugs and the parameters β_1 and β_2 are thus taken to be greater than 0. In addition, from a practical point of view, the response at the control $d_1 = d_2 = 0$ is assumed to be less than 50% and the intercept parameter β_0 is accordingly taken to be less than 0. Note that, without loss of generality, the model can be expressed in terms of the scaled doses $z_1 = \beta_1 d_1$ and $z_2 = \beta_2 d_2$ as

$$\text{logit}(p) = \beta_0 + z_1 + z_2 \text{ with } z_1 \geq 0 \text{ and } z_2 \geq 0. \tag{1}$$

Then the information matrix for the parameters $\beta = (\beta_0, \beta_1, \beta_2)$ at a single observation $z = (z_1, z_2)$ is given by

$$M(\beta; z) = g(z)g(z)^T = \frac{e^u}{(1+e^u)^2} \begin{bmatrix} 1 & z_1 & z_2 \\ z_1 & z_1^2 & z_1 z_2 \\ z_2 & z_1 z_2 & z_2^2 \end{bmatrix}$$

where $g(z) = \dfrac{e^{\frac{u}{2}}}{(1+e^u)}(1, z_1, z_2)$ and $u = \beta_0 + z_1 + z_2$.

Consider now an approximate design which puts weights w_i on the distinct points $z_i = (z_{1i}, z_{2i})$ for $i = 1, \ldots r$, expressed as

$$\xi = \left\{ \begin{array}{ccc} (z_{11}, z_{21}), & \ldots, & (z_{1r}, z_{2r}) \\ w_1, & \ldots, & w_r \end{array} \right\} \quad \text{where} \;\; 0 < w_i < 1 \;\; \text{and} \;\; \sum_{i=1}^{r} w_i = 1.$$

Then the attendant information matrix for the parameters β at the design ξ is given by $M(\beta;\xi) = \sum_{i=1}^{r} w_i g(z_i) g(z_i)^T$. In the present study locally D-optimal designs, that is designs which maximize the determinant of the information matrix at best guesses of the unknown parameters β_0, β_1 and β_2, are sought [Chernoff (1953)].

3 D-optimal designs

3.1 Designs based on 4 points

Consider a 4-point design denoted by $\xi_f^\star$ and given by

$$\xi_f^\star = \left\{ \begin{array}{cccc} (-u-\beta_0,0) & (0,-u-\beta_0) & (u-\beta_0,0) & (0,u-\beta_0) \\ w & w & \frac{1}{2}-w & \frac{1}{2}-w \end{array} \right\}$$

with $0 < u \leq -\beta_0$. The support points lie on the boundary of the design space on lines of constant, complementary u-values and the allocation of the weights is based on symmetry arguments. Note that the constraint on u ensures that the doses are positive. The determinant of the associated information matrix is given by

$$|M(\beta;\xi_f^\star)| = \frac{2e^{3u}u^2 w(1-2w)\{(u-\beta_0)^2 + 8\beta_0 uw\}}{(1+e^u)^6}$$

and is maximized by setting its derivatives with respect to w and u to zero and solving the resultant equations simultaneously. Specifically, the optimal weight satisfies the quadratic equation

$$48\beta_0 uw^2 + 4(u^2 - 6\beta_0 u + \beta_0^2)w - (u-\beta_0)^2 = 0$$

together with the feasibility constraint $0 < w < \frac{1}{2}$ and is given uniquely by

$$w^\star = \frac{-u^2 + 6u\beta_0 - \beta_0^2 + \sqrt{u^2 + 14\beta_0 u + \beta_0^2}}{24\beta_0 u}.$$

It then follows that the optimal u value, denoted by $u^\star$, satisfies the transcendental equation

$$u^2(3+3e^u+2u-2ue^u) + \beta_0^2(1+e^u+2u-2ue^u) + a(1+e^u+u-ue^u) = 0 \quad (2)$$

where $a = \sqrt{u^4 + 14\beta_0^2 u^2 + \beta_0^4}$, together with the constraint $0 < u \leq -\beta_0$. Equation (2) cannot be solved explicitly, only numerically, but it is nevertheless instructive to examine the dependence of the optimal values of u and w on β_0. Values for $u^\star$ and $w^\star$ for selected values of β_0 are presented in Table 1. Note that $u^\star$ decreases monotonically with β_0, that for a value of $\beta_0 = -10$

Table 1. Values of $u^\star$ and $w^\star$ for selected β_0 for 4-point designs

β_0	-5	-4.5	-4	-3.5	-3	-2.5	-2	-1.55
$u^\star$	1.292	1.306	1.323	1.346	1.376	1.418	1.474	1.542
$w^\star$	0.1975	0.1934	0.1888	0.1838	0.1785	0.1731	0.1686	0.1667

the probability of a success at the control $d_1 = d_2 = 0$ is very small (of the order of 4.5×10^{-5}) and that there is a cut-off value of β_0, approximately equal to -1.5434, above which the optimal doses $-u^\star - \beta_0$ become negative. This latter result is discussed in more detail in Sect. 3.3.

The global optimality or otherwise of the proposed D-optimal designs can be confirmed by invoking the appropriate Equivalence Theorem (see [Atkinson and Donev (1992)]) and, specifically, by proving that the directional derivative of the log of the determinant $|M(\beta;\xi)|$ at $\xi_f^\star$ in the direction of $z = (z_1, z_2)$, written $\phi(\xi_f^\star, z, \beta)$, is greater than or equal to 0 over the design space. In fact 4-point designs of the form $\xi_f^\star$ were shown to be globally D-optimal *numerically* for a wide range of β_0 values less than -1.5434. As an example, consider $\beta_0 = -4$. The proposed D-optimal design is given by

$$\xi_f^\star = \left\{ \begin{matrix} (2.677, 0) & (0, 2.677) & (5.323, 0) & (0, 5.323) \\ 0.1888 & 0.1888 & 0.3112 & 0.3112 \end{matrix} \right\}$$

and the directional derivative by

$$\phi(\xi_f^\star, z, \beta) = 3 - \frac{3.955e^{-4+z_1+z_2}(18.817 - 8z_1 - 8z_2 + z_1^2 + z_2^2 + 1.701z_1z_2)}{(1 + e^{-4+z_1+z_2})^2}.$$

A careful search of the values of $\phi(\xi_f^\star, z, \beta)$ over a fine grid of points $z = (z_1, z_2)$ in the region $[0, 10] \times [0, 10]$ indicated that the design $\xi_f^\star$ is indeed globally D-optimal and the 3-dimensional plot of $\phi(\xi_f^\star, z, \beta)$ against $z_1 \geq 0$ and $z_2 \geq 0$ given in Figure 1(a) illustrates this finding. An algebraic proof of the global D-

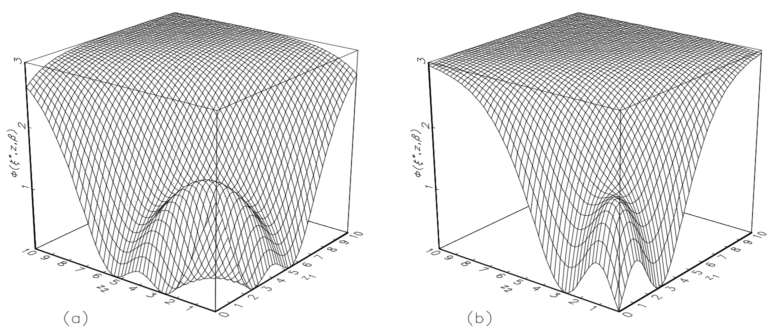

Fig. 1. Plots of the directional derivative $\phi(\xi, z, \beta)$ against z_1 and z_2 for model (1) with (a) $\beta_0 = -4$ and (b) $\beta_0 = -1$.

optimality or otherwise of the proposed 4-point designs was somewhat elusive, the main problems being that the weights assigned to the support points are not equal and that the optimal u value cannot be determined explicitly. A strategy for the required proof is indicated later in the paper.

3.2 Designs based on 3 points

For values of $\beta_0 \geq -1.5434$, the 4-point designs described in the previous section are no longer feasible and it is appealing to consider candidate D-optimal designs which put equal weights on the three support points $(0, 0)$, $(u-$

$\beta_0, 0)$ and $(0, u - \beta_0)$ where $u > \beta_0$. The determinant of the standardized information matrix for the parameters β at such a 3-point design, denoted by $\xi_t^\star$, is given by

$$|M(\beta; \xi_t^\star)| = \frac{(u - \beta_0)^4 e^{\beta_0 + 2u}}{27(1 + e^{\beta_0})^2 (1 + e^u)^4}$$

and the value of u maximizing this determinant satisfies

$$\frac{\partial\,|M(\beta; \xi_t^\star)|}{\partial u} = \frac{(u - \beta_0)^3 (2 - \beta_0 + 2e^u + \beta_0 e^u + u - ue^u)}{27(1 + e^{\beta_0})^2 (1 + e^u)^5} = 0.$$

The solution $u = \beta_0$ is not meaningful since the resultant design comprises the single point $(0, 0)$. Thus the value of u for which $|M(\beta; \xi_t^\star)|$ is a maximum satisfies the equation

$$2 - \beta_0 + 2e^u + \beta_0 e^u + u - ue^u = 0 \tag{3}$$

Numerical studies indicate that there is a unique solution to (3) for values of $u > \beta_0$, say $u^\star$, but this solution does not have an explicit form. Values of $u^\star$ for selected values of β_0 are presented in Table 2.

Table 2. Values of $u^\star$ for selected β_0 for 3-point designs

β_0	-1.5434	-1.5	-1.25	-1	-0.75	-0.5	-0.25	0
$u^\star$	1.5434	1.562	1.674	1.796	1.930	2.075	2.231	2.399

The global D-optimality or otherwise of the candidate designs can be confirmed by demonstrating that the directional derivative $\phi(\xi_t^\star, z, \beta)$ is greater than or equal to zero for all points z in the positive quadrant. This check was performed numerically for selected values of β_0 in the range -1.5434 to 0 using a fine grid of points in the region $[0, 10] \times [0, 10]$ as outlined for the 4-point designs of the previous section. For example, consider $\beta_0 = -1$. The proposed 3-point D-optimal design puts equal weights on the points $(0, 0), (0, 2.796)$ and $(2.796, 0)$, the directional derivative is given by

$$\phi(\xi_t^\star, z, \beta) = 3 - \frac{5.095 e^{-1 + z_1 + z_2} \{2.995 - 2.142(z_1 + z_2) + 0.766 z_1 z_2 + z_1^2 + z_2^2\}}{(1 + e^{-1 + z_1 + z_2})^2}$$

and the 3-dimensional plot of $\phi(\xi_t^\star, z, \beta)$ against $z_1 \geq 0$ and $z_2 \geq 0$ shown in Figure 1(b) indicates that the design is indeed globally optimal. For $\beta_0 \geq -1.5434$, confirming the global optimality or otherwise of 3-point designs of the form $\xi_t^\star$ algebraically is not straightforward however, in particular since the support points of the proposed designs are not associated with complementary u values.

3.3 A special case

The 4-point design introduced in Sect. 3.1 with optimal u value, $u^\star = -\beta_0$, reduces to the 3-point design which puts equal weights on the support points $(0,0), (-2\beta_0^\star, 0)$ and $(0, -2\beta_0^\star)$ where $\beta_0^\star$ satisfies the equation

$$1 + \beta_0 + e^{\beta_0} - \beta_0 e^{\beta_0} = 0 \tag{4}$$

for $\beta_0 < 0$. In other words $\beta_0^\star \approx -1.5434$ and the 3-point design of interest, denoted by $\xi_g^\star$, is given by $(0,0), (3.0868, 0)$ and $(0, 3.0868)$. Note that the support points are associated with the complementary u values, $\pm\beta_0^\star$. The design with $u^\star = -\beta_0 = -\beta_0^\star$ can be shown to be globally D-optimal as follows.

Theorem 1. *Consider the logistic regression model in two variables defined by (1) with $u^\star = \beta_0 = \beta_0^\star$. Then the 3-point design $\xi_g^\star$ which puts equal weights on the support points $(0,0), (-2\beta_0^\star, 0)$ and $(0, -2\beta_0^\star)$ is globally D-optimal.*

Proof. Assume that $\beta_0 = \beta_0^\star$. Then the directional derivative of $\ln |M(\beta;\xi)|$ at $\xi_g^\star$ in the direction of a single point $z = (z_1, z_2)$ is given by

$$\phi(\xi_g^\star, z, \beta) = 3 - 3\frac{e^{\beta_0 + z1 + z2}\ (1 + e^{\beta_0})^2}{e^{\beta_0}\ (1 + e^{\beta_0 + z1 + z2})^2} \left\{ \frac{2\beta_0^2 + z_1^2 + z_2^2 + z_1 z_2 + 2\beta_0(z_1 + z_2)}{2\ \beta_0^2} \right\}.$$

Further, since $u_1 = \beta_0 + z_1 + z_2$ implies $z_2 = u_1 - \beta_0 - z_1$, the directional derivative can be reexpressed as

$$\phi(\xi_g^\star, z, \beta) = 3 - 3\ \frac{e^{u1}\ (1 + e^{\beta_0})^2}{e^{\beta_0}\ (1 + e^{u_1})^2} \left\{ \frac{\beta_0^2 + u_1^2 + (\beta_0 - u_1) z_1 + z_1^2}{2\ \beta_0^2} \right\} \tag{5}$$

with $0 \le z_1 \le u_1 - \beta_0$. It now follows from the Equivalence Theorem for D-optimal designs that the design $\xi_g^\star$ is globally D-optimal provided the condition $\phi(\xi_g^\star, z, \beta) \ge 0$ holds. Consider u_1 fixed, i.e. consider points z on a line of constant logit. Then $\phi(\xi_g^\star, z, \beta)$ given by (5) is proportional to the quadratic function $f(z_1) = \beta_0^2 + u_1^2 + z_1(\beta_0 - u_1) + z_1^2$ which has a unique minimum at $z_1 = \frac{u_1 - \beta_0}{2}$. Therefore, the maxima of $f(z_1)$ within the design space are located at the boundary points $z_1 = 0$ and $z_1 = u_1 - \beta_0$. Thus the minima of the directional derivative $\phi(\xi_g^\star, z, \beta)$ for all points z in the positive quadrant occur on the boundaries $z_1 = 0$ and $z_2 = 0$. Now on setting $z_1 = 0$ or $z_1 = u_1 - \beta_0$ in (5), the inequality $\phi(\xi_g^\star, z, \beta) \ge 0$ reduces to

$$2\frac{e^{\beta_0}\ (1 + e^{u_1})^2}{e^{u_1}\ (1 + e^{\beta_0})^2} \ge \frac{(\beta_0^2 + u_1^2)}{\beta_0^2}. \tag{6}$$

This condition, together with the fact that β_0 satisfies equation (4) and thus $\beta_0 = \beta_0^\star$, is precisely the condition which emerges in invoking the appropriate directional derivative to prove the global optimality of the D-optimal design for a logistic regression model with one explanatory variable. Thus it follows immediately from that setting that condition (6) holds for all $u_1 \in I\!R$ and thus, in the present case, for all feasible $u_1 \ge \beta_0^\star$.

The framework of the above theorem can be used to devise a strategy for proving the global D-optimality of the candidate 3- and 4-point designs discussed in the earlier sections.

4 Conclusions

The main aim of the present study has been to construct locally D-optimal designs for the logistic regression model in two variables subject to the constraint that the values of the variables are greater than or equal to zero. In particular it is shown that the designs so constructed depend on the parameters β_1 and β_2 of model (1) only through the scaling of the two explanatory variables but that the basic pattern of the designs is determined by the intercept parameter β_0. Specifically, if $\beta_0 < \beta_0^\star$ where $\beta_0^\star$ satisfies equation (4) then the D-optimal design is based on 4 points of support located on complementary logit lines, whereas if $\beta_0^\star \leq \beta_0 \leq 0$ then the design comprises 3 points including a control. The global D-optimality of the designs for a wide range of β_0 values was demonstrated numerically but was only proved algebraically for the case with $\beta_0 = \beta_0^\star$. The broad strategy used in the proof for the latter setting, that is in Theorem 1, should be applicable to all D-optimal designs reported here. However the extension is not entirely straightforward and is currently being investigated.

There is much scope for further work. In particular there is a need to relate the D-optimal designs constructed here to the geometry of the design locus as elucidated in [Sitter and Torsney (1995)], to the "minimal" point designs developed in [Torsney and Gunduz (2001)] and to the results of [Wang et al. (2006)] on Poisson regression. In fact the work reported here forms part of a larger study aimed at identifying patterns and taxonomies of designs for the logistic regression model in two variables both with and without an interaction term and with a range of constraints on the variables. Finally it would be interesting to extend the design construction to accommodate other criteria and, following [Torsney and Gunduz (2001)], to logistic regression models with more than two explanatory variables.

Acknowledgements: The authors would like to thank the University of Cape Town, the University of KwaZulu-Natal and the National Research Foundation, South Africa, for financial support.

References

[Atkinson (2006)] Atkinson, A.C.: Generalized linear models and response transformation. In: Khuri A.I. (ed) Response Surface Methodology and Related Topics. World Scientific, New Jersey (2006)

[Atkinson and Donev (1992)] Atkinson, A.C., Donev, A.N.: Optimum Experimental Designs. Clarendon Press, Oxford (1992)

[Atkinson and Haines (1996)] Atkinson, A.C., Haines L.M.: Designs for nonlinear and generalized linear models. In: Ghosh S., Rao C.R. (ed) Handbook of Statistics, Volume 13. Elsevier, Amsterdam (1996)

[Chernoff (1953)] Chernoff, H.: Locally optimal designs for estimating parameters. Ann. Math. Statist., **24**, 586–602 (1953)

[Jia and Myers (2001)] Jia, Y., Myers R.H.: Optimal experimental designs for two-variable logistic regression models. Preprint (2001)

[Kupchak (2000)] Kupchak P.I. : Optimal designs for the detection of drug interaction. Ph.D. Thesis, University of Toronto, Toronto (2000)

[Sitter and Torsney (1995)] Sitter R.R., Torsney B.: Optimal designs for binary response experiments with two design variables. Statistica Sinica, **5**, 405–419 (1995)

[Torsney and Gunduz (2001)] Torsney B., Gunduz N.: On optimal designs for high dimensional binary regression models. In: Atkinson A.C., Bogacka B., Zhigljavsky A. (ed) Optimum Design 2000. Kluwer, Dordrecht (2001)

[Wang et al. (2006)] Wang Y., Myers R.H., Smith E.P., Ye K.: *D*-optimal designs for Poisson regression models. J. Statist. Plann. Inference, **136**, 2831–2845 (2006)