ARTICLE IN PRESS

Journal of Statistical Planning and Inference ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

D-optimal designs for the two-variable binary logistic regression model with interaction

Linda M. Haines [a,*], Gaëtan M. Kabera [b], Principal Ndlovu [b]

[a] *Department of Statistical Sciences, University of Cape Town, Private Bag X3, Rondebosch 7701, South Africa*
[b] *Department of Statistics, University of South Africa, South Africa*

ARTICLE INFO

Article history:
Received 31 March 2017
Received in revised form 24 August 2017
Accepted 26 August 2017
Available online xxxx

Keywords:
D-optimality
Binary logistic model
Interaction
Proof-of-concept

ABSTRACT

The analytical construction of optimal designs for generalized linear models with interaction between the explanatory variables is challenging and the optimal designs reported in the statistical literature for such models are invariably constructed numerically. In the present paper, the two-variable binary logistic model with interaction is introduced and the problem of constructing approximate globally *D*-optimal designs for the model within the context of drug combination studies is considered. The requisite designs are found by a blend of analytical and numeric techniques and are shown to depend sensitively on the value of the intercept parameter. More specifically, for settings in which the intercept parameter is greater than or equal to a specified bound, the globally *D*-optimal designs are shown to be based on four equally-weighted points. In contrast, for those settings in which the intercept parameter is less than the specified bound, the globally *D*-optimal designs are shown to comprise either four, five or six points, with only the 5-point designs obtained explicitly. The results of this study are illustrated throughout by means of examples involving a range of parameter settings. In addition, a real world example is introduced and the practical advantages which accrue from implementing exact designs based on approximate *D*-optimal and near-*D*-optimal designs are highlighted.

1. Introduction

Designs for binary logistic models which are linear or quadratic in two or more variables are well-documented and comprehensively reviewed in the article by Atkinson and Woods (2015). Specifically, if the logit depends linearly on the explanatory variables, constraints must be placed on the design space in order to develop explicit results. Results in this context are reported in, for example, Sitter and Torsney (1995), Yang et al. (2011) and Kabera et al. (2012). Models for which the logit depends on a full quadratic in two variables arise in the context of response surface methodology and have been investigated numerically in Atkinson et al. (2007) and Atkinson and Woods (2015). However few studies have been devoted to designs for binary logistic models in two or more variables which incorporate an interaction term. Kupchak (2000), in his Ph.D. thesis, explored a range of designs for the precise estimation of the parameters of the two-variable model using a blend of theory and calculation, while Jia and Myers (2001) adopted a geometric approach in an attempt to address the design problem analytically. More recently, Woods et al. (2006), Dror and Steinberg (2006) and McGree and Eccleston (2008) report optimal designs for the two-variable binary logistic model with interaction which were constructed numerically and are introduced as examples in a broader computational context. It would seem however that analytic studies on the designs

* Corresponding author.
E-mail address: linda.haines@uct.ac.za (L.M. Haines).

http://dx.doi.org/10.1016/j.jspi.2017.08.007
0378-3758/

Please cite this article in press as: Haines L.M., et al., *D*-optimal designs for the two-variable binary logistic regression model with interaction. J. Statist. Plann. Inference (2017), http://dx.doi.org/10.1016/j.jspi.2017.08.007.

are challenging to conduct and remain largely unreported in the literature. Indeed, Yang et al. (2011) state that "[it] is not clear whether there exists a general structure for this type of optimal design".

Drug combination studies to identify synergy, non-interaction or antagonism between two drugs are conducted extensively in modern medicine. The statistical models most commonly used to analyze data from such studies are reviewed in the paper by Greco et al. (1995) and more succinctly in the recent papers by Lee et al. (2007) and Harbron (2010). In particular, the two-variable logistic model with interaction in which the response is binary was introduced into drug combination studies by Carter et al. (1988) and is classified as a response surface model within that context. The model yields a global assessment of the action of the drugs and, in addition, a meaningful interaction index and is widely used in practice (Lee et al., 2007). However the problem of constructing designs for binary logistic models with interaction which are, in some sense, optimal does not appear to have been addressed in the drug combination literature.

The aim of the present paper is to develop a structure for *D*-optimal designs for two-variable binary logistic models with interaction within the context of drug combination studies by using a mix of algebra and computation. Certain constraints are placed on the model and on the design setting however. Thus, it is assumed that the explanatory variables relate to the concentration of two drugs and are greater than or equal to zero, that the placebo effect is less than 50%, that both drugs have a positive effect on the binary response and that the interaction between the drugs is synergistic.

The paper is structured as follows. Some preliminary results are presented in Section 2. Four-point designs, which are globally *D*-optimal under a specified condition on the intercept parameter, are then introduced in Section 3 and *D*-optimal designs for settings in which the specified condition does not hold are formulated and examined in Section 4. A real world example is introduced in Section 5 and some conclusions and pointers for further research are given in Section 6.

2. Preliminaries

Consider a binary response variable Y which follows a Bernoulli distribution and takes the values one for a success or positive response and zero for a failure or negative response. If Y is related to two predictor or explanatory variables x_1 and x_2 and to their interaction x_1x_2 through the two-variable binary logistic model, then the probability of success, p, can be expressed in terms of the logit

$$u = \text{logit}(p) = \ln\frac{p}{1-p} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 \tag{1}$$

where β_0, β_1, β_2 and β_{12} are unknown parameters.

In the present study, the variables x_1 and x_2 are taken to be concentrations of the doses of two drugs with $x_1 \geq 0$ and $x_2 \geq 0$. Certain assumptions relating to the parameters are also introduced within this context and are used consistently throughout the text. Specifically, when an (x_1, x_2)-dose space with $x_1 \geq 0$ and $x_2 \geq 0$ is used, the placebo effect with $x_1 = x_2 = 0$ is expected to be less than 50%, that is $p < 0.5$, and thus the intercept parameter β_0 is taken to be negative (Kabera et al., 2015). In addition, the probability of a positive response is expected to increase with dose concentrations for both drugs and thus β_1 and β_2 are taken to be greater than zero. Finally, the parameter β_{12} represents the interaction effect between the two drugs, with non-interaction or additivity, synergy and antagonism between the drugs observed when $\beta_{12} = 0$, $\beta_{12} > 0$ and $\beta_{12} < 0$, respectively (Kabera, 2009). Synergy is assumed to hold throughout the paper, that is $\beta_{12} > 0$. Results related to those reported here but for antagonism can be derived in a similar manner (Kabera, 2009), while those for additivity, that is for the logit model $u = \beta_0 + \beta_1 x_1 + \beta_2 x_2$, are presented in Kabera et al. (2015).

The lines of constant logit are hyperbolae centered at $(-\frac{\beta_2}{\beta_{12}}, -\frac{\beta_1}{\beta_{12}})$. It is instructive, following Jia and Myers (2001), to examine the logit pairs $u > 0$ and $-u < 0$ for the setting $\beta_0 < 0$, $\beta_1 > 0$, $\beta_2 > 0$ and $\beta_{12} > 0$ geometrically. Specifically, the pairs of logit lines $\pm u$ which intersect the axes at $x_1, x_2 > 0$ for the representative cases $u > -\beta_0^\star > -\beta_0 > 0$, $-\beta_0^\star > u > -\beta_0 > 0$ and $-\beta_0^\star > -\beta_0 > u > 0$ where $\beta_0^\star = \beta_0 - \frac{\beta_1\beta_2}{\beta_{12}}$ are shown in Fig. 1(a), (b) and (c), respectively. It is relevant here to note that for $u > -\beta_0 > 0$ only the logit line u intersects the axes at $x_1, x_2 > 0$ and that otherwise, for $0 < u < -\beta_0$, both the logit lines $\pm u$ intersect the axes at $x_1, x_2 > 0$.

Consider now an approximate design, often termed a continuous design Atkinson et al. (2007, p. 119), for the logit model (1) defined by

$$\xi = \begin{Bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \dots & \boldsymbol{x}_d \\ w_1 & w_2 & \dots & w_d \end{Bmatrix}$$

where $\boldsymbol{x}_i = (x_{i1}, x_{i2})$ are the support points in the domain $\mathcal{D} = [0, \infty) \times [0, \infty)$ and w_i are the design weights satisfying the conditions $0 < w_i < 1$, $\sum_{i=1}^{d} w_i = 1$ for $i = 1, \dots, d$ with $d \geq 4$. Then the information matrix for the parameter vector $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_{12})^T$ at a single design point $\boldsymbol{x} = (x_1, x_2)$ is given by

$$M(\boldsymbol{x}; \boldsymbol{\beta}) = \frac{e^u}{(1+e^u)^2}\widetilde{\boldsymbol{x}}\widetilde{\boldsymbol{x}}^T$$

where $\widetilde{\boldsymbol{x}} = (1, x_1, x_2, x_1x_2)^T$. The information matrix evaluated at the approximate design follows immediately as

$$M(\xi; \boldsymbol{\beta}) = \sum_{i=1}^{d} w_i \frac{e^{u_i}}{(1+e^{u_i})^2}\widetilde{\boldsymbol{x}}_i\widetilde{\boldsymbol{x}}_i^T$$

Please cite this article in press as: Haines L.M., et al., *D*-optimal designs for the two-variable binary logistic regression model with interaction. J. Statist. Plann. Inference (2017), http://dx.doi.org/10.1016/j.jspi.2017.08.007.