Computational Statistics and Data Analysis 71 (2014) 1066–1076

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Bayesian *D*-optimal designs for the two parameter logistic mixed effects model

Haftom T. Abebe*, Frans E.S. Tan, Gerard J.P. Van Breukelen, Martijn P.F. Berger

Maastricht University, Department of Methodology and Statistics, P.O. Box 616, 6200 MD, Maastricht, The Netherlands

ARTICLE INFO

Article history:
Received 26 April 2012
Received in revised form 29 July 2013
Accepted 30 July 2013
Available online 16 August 2013

Keywords:
Bayesian *D*-optimal designs
Logistic mixed effects model
Maximin Bayesian *D*-optimal design
Maximin locally *D*-optimal designs
Minimum relative efficiency

ABSTRACT

Bayesian optimal designs for binary longitudinal responses analyzed with mixed logistic regression describing a linear time effect are considered. In order to find the optimal number and allocations of time points, for different priors, cost constraints and covariance structures of the random effects, a scalar function of the approximate information matrix based on the first order penalized quasi likelihood (PQL1) is optimized. To overcome the problem of dependence of Bayesian designs on the choice of prior distributions, maximin Bayesian *D*-optimal designs are proposed. The results show that the optimal number of time points depends on the subject-to-measurement cost ratio and increases with the cost ratio. Furthermore, maximin Bayesian *D*-optimal designs are highly efficient and robust under changes in priors. Locally *D*-optimal designs are also investigated and maximin locally *D*-optimal designs are found to have much lower minimum relative efficiency and fewer time points than maximin Bayesian *D*-optimal designs. When comparing the efficiencies of designs with equidistant time points with the Bayesian *D*-optimal designs, it was found that three or four equidistant time points are advisable for small cost ratios and five or six equidistant time points for large cost ratios.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Over the past few decades the design of longitudinal studies for dichotomous data has not received much attention from statisticians. To a large extent the precision of estimators of the unknown parameters of the statistical model depends on the design used in the experiment (Khuri et al., 2006). A researcher designing a longitudinal study has to decide on the number and allocation of time points. Optimal design theory can help in making this decision. However, optimal design depends on the unknown parameter values in non-linear models like the logistic regression model. Much effort has been devoted to this problem. A common approach is to design an experiment to be efficient for a best guess of the parameter values, which leads to locally optimal designs (see, e.g. Chernoff, 1953; Abdelbasit and Plackett, 1983). Such a design, however, may not be efficient for other parameter values. Thus, the ability to find locally optimal designs still faces the serious problem of how to take into account the uncertainty about the parameter values.

To handle the problem of local optimality, three methods are proposed in the literature namely, the sequential procedure (see, e.g. Abdelbasit and Plackett, 1983; Wu, 1985), the minimax or maximin approach (see, e.g. Dette, 1997; Ouwens et al., 2002; Tekle et al., 2008; Tan, 2011), and the use of Bayesian methods, described by Chaloner and Larntz (1989), Chaloner and Verdinelli (1995), Atkinson et al. (2007), Tommasi and López-Fidalgo (2010), and Drovandi et al. (2013). Due to restrictions

* Correspondence to: Peter Debyeplein 1 6229 HA Maastricht, P.O. BOX 616, 6200 MD Maastricht, The Netherlands. Tel.: +31 433882905; fax: +31 433618388.

E-mail addresses: haftom.temesgen@maastrichtuniversity.nl, haftomd@yahoo.com (H.T. Abebe).

0167-9473/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.csda.2013.07.040

on the length of the research period, the sequential approximation of an optimal design may not be feasible. A potential objection to the maximin approach is that the maximin design may occur at the edges of the parameter space and, if some points in the parameter space have a very low probability, then their weight may be over-emphasized (Atkinson et al., 2007).

In the present paper, we will use the Bayesian approach, which is a more natural approach that formally accounts for the prior uncertainty of the parameter values. The Bayesian design literature outside the normal linear model is mostly restricted to binary response models (see, e.g. Khuri et al., 2006). Optimal designs for fixed effects models for binary data, especially logistic models have been investigated in the literature. Optimal designs were discussed by Bischoff (1993), Mentrē et al. (1997), Atkins and Cheng (1999), Tan and Berger (1999), Berger and Tan (2004) and Berger and Wong (2009) for the linear random effects model. Recently, Ouwens et al. (2006) and Tekle et al. (2008) have extended the work on optimal designs for logistic models with random effects using maximin D-optimal designs. However, to the authors' knowledge no work has been done on Bayesian optimal designs for longitudinal studies with a binary response. In this paper, we will investigate numerically Bayesian D-optimal designs for the two parameter logistic mixed effects model with repeated measures of a binary response.

The paper is organized as follows. The next section introduces the logistic mixed effects model and the variance–covariance matrix of parameter estimators. In Section 3 the Bayesian D-optimality criterion and the corresponding algorithm to construct these designs are given as well as a discussion of a linear cost function and the relative efficiency as a measure for the comparison of designs. In Section 4 a numerical study and its results are presented, followed by the conclusions in Section 5.

2. Logistic mixed effects models

Let the $q \times 1$ vector $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{iq})'$ be the binary responses y_{ij} (either 0 or 1) of subject i at q time points, $i = 1, 2, \ldots, N$ and $j = 1, \ldots, q$. It is assumed that, conditional on the subject-specific random effect vector $\boldsymbol{b}_i$, the binary responses y_{ij} are independently Bernoulli distributed with probability of success $p(y_{ij} = 1|\boldsymbol{b}_i)$. These probabilities are related to the fixed and random-effects via the logit link. The corresponding logistic mixed effects model is given by:

$$\operatorname{logit}(p(y_{ij} = 1|\boldsymbol{b}_i)) = \log\left(\frac{p(y_{ij} = 1|\boldsymbol{b}_i)}{1 - p(y_{ij} = 1|\boldsymbol{b}_i)}\right) = \boldsymbol{x}_j'\boldsymbol{\beta} + \boldsymbol{z}_j'\boldsymbol{b}_i, \tag{1}$$

where the $p \times 1$ vector $\boldsymbol{x}_j$ is the design vector of the explanatory variables at the jth measurement for subject i, $\boldsymbol{\beta}$ is the corresponding $p \times 1$ vector of unknown fixed effects, and $\boldsymbol{z}_j$ is the $r \times 1$ design vector for the random effects that is usually a subset of vector $\boldsymbol{x}_j$, $\boldsymbol{b}_i$ is the corresponding $r \times 1$ vector of unknown random effects, which is assumed to have a multivariate normal distribution with mean zero and covariance matrix $\boldsymbol{D}$.

To simplify, let us assume that all subjects have measurements at the same time points, but this can be relaxed. In this paper, we consider a linear ($p = 2$) time effect, i.e., the design vector is $\boldsymbol{x}_j' = (1\ t_j)$ and $\boldsymbol{\beta} = (\beta_0\ \beta_1)'$, where t_j is the time point of the jth measurement, and β_0 and β_1 are the fixed intercept and slope effects. Suppose that for the linear time effect, a random intercept and random slope are assumed. The design vector is then $\boldsymbol{x}_j' = \boldsymbol{z}_j' = (1\ t_j)$ and $\boldsymbol{b}_i = (b_{0i}\ b_{1i})'$, where b_{0i} and b_{1i} are the corresponding random (subject-specific) deviations from these fixed effects. Thus, according to model (1), the log-odds of a positive response for subject i at time t_j is given by:

$$\operatorname{logit}(p(y_{ij} = 1|\boldsymbol{b}_i)) = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t_j. \tag{2}$$

The objective of this paper is to determine the optimal number of time points q and the optimal allocation of these time points t_j ($j = 1, \ldots, q$), where 'optimal' means that the parameters $\boldsymbol{\beta}$ are estimated as efficiently as possible.

2.1. Variance–covariance matrix of the parameter estimators

The log-likelihood cannot be written down in closed form due to the random effects in model (1). Hence, either numerical methods or approximations to the log-likelihood must be used. Numerical methods require large computational resources and more importantly they require full knowledge of the data (Moerbeek et al., 2003; Han and Chaloner, 2004; Tekle et al., 2008), making them computationally inconvenient for optimal design procedures. Therefore, we will use an approximate information matrix based on a first order penalized quasi-likelihood (PQL1). See Breslow and Clayton (1993), and Jang and Lim (2009) for details. PQL1 performs best in terms of point estimation since it produces the smallest mean squared error (MSE) and the bias of the estimators decreases as the sample size increases (Breslow and Clayton, 1993; Moerbeek et al., 2003; Jang and Lim, 2009). The asymptotic variance–covariance matrix of $\hat{\boldsymbol{\beta}}$, which is the inverse of the information matrix of the parameters, is approximated in PQL1 by:

$$\operatorname{var}(\hat{\boldsymbol{\beta}}) \approx (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}, \tag{3}$$

where $\hat{\boldsymbol{\beta}}$ is the estimator of $\boldsymbol{\beta}$ for the logistic mixed-effect model (1), $\boldsymbol{X}$ is the $Nq \times p$ design matrix formed by stacking $\{\boldsymbol{x}_j'\}$ for N subjects, and the $Nq \times Nq$ block-diagonal matrix $\boldsymbol{V}$ has N blocks of $q \times q$ variance–covariance matrices given by:

$$\boldsymbol{v}_i \approx \boldsymbol{w}_i^{-1} + \boldsymbol{Z}\,\boldsymbol{D}\,\boldsymbol{Z}', \tag{4}$$

Download English Version:

https://daneshyari.com/en/article/6870619

Download Persian Version:

https://daneshyari.com/article/6870619

Daneshyari.com