Available online at www.sciencedirect.com
ScienceDirect
Journal of Statistical Planning and Inference 137 (2007) 3882–3893
journal of statistical planning and inference
www.elsevier.com/locate/jspi

Design optimality in multi-factor generalized linear models in the presence of an unrestricted quantitative factor

Ulrike Graßhoff[a], Heiko Großmann[b], Heinz Holling[c], Rainer Schwabe[a,*]

[a] *Institute for Mathematical Stochastics, Otto von Guericke University, PF 4120, D-39 016 Magdeburg, Germany*
[b] *School of Mathematical Sciences, Queen Mary, University of London, Mile End Road, London E1 4NS, UK*
[c] *Psychologisches Institut IV, Westfälische Wilhelms-Universität, Fliednerstr. 21, D-48149 Münster, Germany*

Received 29 December 2005; received in revised form 7 September 2006; accepted 6 April 2007
Available online 4 May 2007

Abstract

In this paper we show that product type designs are optimal in partially heteroscedastic multi-factor linear models. This result is applied to obtain locally *D*-optimal designs in multi-factor generalized linear models by means of a canonical transformation. As a consequence we can construct optimal designs for direct logistic response as well as for Bradley–Terry type paired comparison experiments.

MSC: 62K05; 62J12; 62J15

Keywords: Optimal design; Logistic model; Generalized linear model; Heteroscedastic linear model; Additive model; Product type design; Paired comparison; Bradley–Terry model

1. Introduction

In applications multi-factor generalized linear models attract growing attention (McCullagh and Nelder, 1989). For these models much research has been done about statistical inference, but there are only few attempts to design such experiments (Ford et al., 1992; Sitter and Torsney, 1995; Atkinson and Haines, 1996). Many of the complications that arise in that context are due to the fact that in contrast to the general linear model the design here depends on the unknown parameter vector of interest. Results obtained within the linear model framework can often be immediately applied to generalized linear models, however, when it is assumed that the parameter vector of interest is equal to zero. In a marketing research context it was for example pointed out by Großmann et al. (2002) that optimal designs for linearizations of generalized linear models are also optimal for the original model under that assumption.

In this paper we demonstrate how designs for the linear model can be used to construct optimal designs for a certain class of generalized linear models without imposing any restrictions on the parameter vector of interest. The results presented here are important since they provide a link between the vast literature on the design of experiments for linear models and generalized linear models which can be exploited beneficially for generating practical designs for the latter class of models. In particular, we demonstrate that the parameter dependency of the design problem can be eliminated

* Corresponding author. Tel.: +49 391 67 18304; +49 391 67 11172.
E-mail address: rainer.schwabe@mathematik.uni-magdeburg.de (R. Schwabe).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2007.04.007

to a certain extent allowing the use of linear model designs as bricks for optimal designs in generalized linear models. The set of factors in the generalized linear models we consider contains a single unrestricted quantitative factor. This enables us to make use of the concept of a canonical transformation (Ford et al., 1992; Sitter and Torsney, 1995) and product type designs (Schwabe, 1996) for generating optimal designs.

The usefulness of this approach is indicated for multi-factor logistic models. We distinguish two cases. First, logistic models in which the response is directly observable are considered. These models have applications, for instance, in educational testing and more generally explanatory item response modelling (De Boeck and Wilson, 2004). Second, we consider Bradley–Terry-type logistic models for paired comparisons. In this situation there are no direct responses to combinations of factor levels available. Rather combinations of levels are presented in pairs and it is recorded which of the alternatives in a pair is chosen. Paired comparisons and more generally choice models with more than two alternatives are widely used in marketing, health economics and other fields to learn about consumer or patient preferences and the design problem for those models has received considerable attention over the past years (for reviews, see Großmann et al., 2002; Louviere et al., 2004).

An essential concept used in the marketing literature on paired comparisons and choice designs is the idea of utility balance (Huber and Zwerina, 1996) which aims at designing the choice alternatives in such a way that they are of equal value for the respondent. In particular, Kanninen (2002) showed that in a logistic response setup alternatives are optimal which have optimal utility difference. Consistent with the setup considered here, these results require that one of the factors is quantitative, i.e. it can be adjusted continuously, and unrestricted, which will be typically the price of the alternatives measured on a logarithmic scale. This approach has attracted interest in applications, as the so obtained optimal designs have, at least, formally a much simpler structure than in purely qualitative layouts (Graßhoff and Schwabe, 2005). Commonly local *D*-optimality is considered which aims at minimizing the determinant of the asymptotic covariance matrix.

In the following we restrict ourselves to the concept of approximate designs (see Silvey, 1980), but which can, here, frequently be realized with sufficiently small sample sizes, at least, after employing suitable reduction principles (see Graßhoff et al., 2004; Burgess and Street, 2005, for the corresponding linear paired comparison model).

In Section 2 we prove the optimality of certain product type designs in partially heteroscedastic linear models as an auxiliary result. Based on these findings optimal designs are obtained in Section 3 for generalized linear models by means of a canonical transformation (Sitter and Torsney, 1995). In Section 4 optimal designs are derived for the case of logistic response, considered in the marketing research literature, both for direct observations and paired comparisons. As a consequence it is established that in the latter case optimal designs are given by an optimal utility difference, i.e. the probabilities of alternatives should be equal to some optimal value depending solely on the number of parameters in the model.

2. Optimal designs for linear additive models

In this section we derive a characterization of optimal designs for additive multi-factor models with heteroscedastic errors. The result serves as a tool for multi-factor generalized linear models but may have some interest on its own. In fact, it generalizes results by Rafajłowicz and Myszka (1992) and Schwabe (1996, Section 5) to heteroscedastic response but avoids the restrictions imposed by Rodríguez and Ortiz (2005).

2.1. Linear models with a constant term

First, we consider the linear additive model

$$Y(\mathbf{x}_1, \mathbf{x}_2) = \beta_0 + \mathbf{f}_1(\mathbf{x}_1)^\top \boldsymbol{\beta}_1 + \mathbf{f}_2(\mathbf{x}_2)^\top \boldsymbol{\beta}_2 + \varepsilon(\mathbf{x}_1, \mathbf{x}_2), \tag{1}$$

$\mathbf{x}_1 \in \mathscr{X}_1$, $\mathbf{x}_2 \in \mathscr{X}_2$, with explicit constant term β_0 and potentially heteroscedastic observations, $\mathrm{Var}(Y(\mathbf{x}_1, \mathbf{x}_2)) = \sigma^2(\mathbf{x}_1, \mathbf{x}_2)$. In the linear setting we denote by $\lambda(\mathbf{x}_1, \mathbf{x}_2) = 1/\sigma^2(\mathbf{x}_1, \mathbf{x}_2)$ the intensity (or efficiency) function (see e.g. Fedorov, 1972, p. 66) which measures the contribution of the setting $(\mathbf{x}_1, \mathbf{x}_2)$ to the experiment. Here, we assume additionally that the variance and, hence, the intensity function

$$\lambda(\mathbf{x}_1, \mathbf{x}_2) = \lambda_2(\mathbf{x}_2) \tag{2}$$

depends on the second component only.